Report No. UT-19.15

BALANCED ASPHALT CONCRETE MIX PERFORMANCE IN UTAH

PHASE III: EVALUATION OF FIELD MATERIALS USING BBR AND SCB-IFIT TESTS

Prepared For:

Utah Department of Transportation
Research and Innovation Division

Submitted By:

University of Utah
Department of Civil and Environmental Engineering

Authored By:

Pedro Romero, Ph.D., P.E.
Kevin VanFrank, P.E.

Final Report
May 2019

RESEARCH
Keeping Utah Moving
Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Scott Andrus
- Howard Anderson
- Steve Anderson
- David Stevens

The authors also acknowledge the following individuals and their work. A significant portion of the information contained in this report is the result of their hard work. This includes the following current and former graduate students at the University of Utah:

- Abu Sufian Mohammad Asib, Ph.D. candidate
- Faramarz Safazadeh, Ph.D. candidate
- Shuangli Bao, M.Sc.

The following people also provided valuable input for this work:

- Tim Biel from PEPG
- Michael VanMilligen from PEPG
- Clark Allen from Utah Department of Transportation

The authors also acknowledge the financial contribution from the Mountains Plains Consortium.

TECHNICAL REPORT ABSTRACT

1. Report No. UT-19.15	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle BALANCED ASPHALT CONCRETE MIX PERFORMANCE IN UTAH, PHASE III: EVALUATION OF FIELD MATERIALS USING BBR AND SCB-IFIT TESTS		5. Report Date May 2019 6. Performing Organization Code N/A
7. Author(s) Pedro Romero, Ph.D., P.E. and Kevin VanFrank, P.E.		8. Performing Organization Report No. N/A
9. Performing Organization Name and Address University of Utah, Dept. of Civil and Environmental Engineering, 110 Central Campus Drive #2000, Salt Lake City, UT 84112 PEPG Consulting, 9270 South 300 West, Suite A-2, Sandy, UT 84070		10. Work Unit No. 5H08012H 11. Contract or Grant No. 17-9021
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410		13. Type of Report & Period Covered Final Report May 2017 to March 2019 14. Sponsoring Agency Code PIC No. UT16.101

15. Supplementary Notes

Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

A study was conducted to evaluate the applicability of two cracking tests for low- and intermediate-temperature performance for the purpose of creating a balanced hot-mix asphalt mix. For the low-temperature performance, the Bending Beam Rheometer (BBR) test on mixtures based on AASHTO TP-125-16 protocols was studied. For intermediate-temperature performance, the Flexibility Index (FI) of hot-mix asphalt as determined using the Semi-circular bend (SCB) configuration based on AASHTO TP-124-16 was studied. Seven asphalt mixtures were collected across the State of Utah at two locations: at the plant and at laydown from the windrow.

For the low-temperature performance, it was determined that the specification proposed as part of a previous study is adequate and will allow evaluation of the low-temperature performance of asphalt mixtures at specific environments. For the intermediate-temperature test, it was found that the variability in the results (FI) continues to be a problem. At least 8 samples should be tested to obtain an average that represents the actual value within 20%. This requires compaction of 2 gyratory pucks.

Regarding the Flexibility Index, asphalt mixtures sampled at the plant can be expected to have an FI generally between 3.0 and 20.0. The material sampled at laydown can be expected to have an FI somewhere between 2.1 and 18.5. A small improvement in the differences between labs was seen when testing was done at a loading rate of 15 mm/min but there was no clear benefit regarding the coefficient of variation. Different loading rates resulted in different performance ranking of the mixtures; however, without any performance information available from the field, it is not known if one loading rate is preferred over the other.

17. Key Words Asphalt Concrete, Pavement Cracking, Bending Beam Rheometer, Flexibility Index, Recycled Asphalt Pavement		18. Distribution Statement Not restricted. Available through: UDOT Research & Innov. Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research		23. Registrant's Seal N/A
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 74	22. Price N/A	

TABLE OF CONTENTS

<u>**LIST OF TABLES**</u>

LIST OF FIGURES

UNIT CONVERSION FACTORS

Units used in this report and not conforming to the UDOT standard unit of measurement (U.S. Customary system) are given below with their U.S. Customary equivalents:

- 25.4 millimeters (mm) = 1 inch (in)
- 1 megapascal (MPa) = 145.04 pounds per square inch (psi)
- 1 Newton (N) = 0.2248 pounds force (lbs)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. (Adapted from FHWA report template, Revised March 2003)

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
ANOVA	Analysis of Variance
APWA	American Public Works Association
ASTM	American Society of Testing and Materials
BBR	Bending Beam Rheometer, refers to AASHTO TP-125-16
CoV	Coefficient of Variation – Percent Ratio of the Standard Deviation to the Mean
DCSE	Dissipated Creep Strain Energy
DCT	Disc-shaped Compact Test
DOT	Department of Transportation
FE	Fracture energy
FHWA	Federal Highway Administration
FI	Flexibility Index, AASHTO TP-124-16
Gmm	Maximum theoretical specific gravity, AASHTO T-209
IDT	Indirect Tensile Test
IFIT	Illinois Flexibility Index Test, AASHTO TP-124-16
MOI	Manual of Instruction
m-value	Slope of the log of the modulus versus the log of time in the BBR test
NDES	Designed number of gyrations
NMAS	Nominal Maximum Aggregate Size
PG	Performance Grade, AASHTO M-320
RAP	Recycled Asphalt Pavement
SCB	Semi-Circular Bending Test
SD	Standard Deviation
SGC	Superpave gyratory Compactor, AASHTO T-312
UDOT	Utah Department of Transportation

EXECUTIVE SUMMARY

A study was conducted to evaluate the applicability of two cracking tests for low- and intermediate-temperature performance for the purpose of creating a balanced hot-mix asphalt mix. For the low-temperature performance, the Bending Beam Rheometer (BBR) test on mixtures based on AASHTO TP-125-16 protocols was studied. For intermediate-temperature performance, the Flexibility Index (FI) of hot-mix asphalt as determined using the Semi-circular bend (SCB) configuration based on AASHTO TP-124-16 was studied. Seven asphalt mixtures, intended for both the state DOT and different municipalities, were collected across the State of Utah at two locations: at the plant from the slats and at laydown from the windrow. These mixtures are considered representative of what is being produced across the State of Utah.

For the low-temperature evaluation, it was confirmed that the specification proposed as part of a previous study is adequate and will allow evaluation of the expected performance of asphalt mixtures at specific low-temperature environments. While most of the mixtures produced have a relatively high creep modulus at the intended environment (creep modulus >12,000 MPa), their relatively high relaxation capacity (m-value >0.12) should result in good performance. These predictions are based on the mixture as a system and are not based on individual parameters such as neat asphalt binder grade or Recycled Asphalt Pavement (RAP) content. It was also determined that the aging that occurs between the plant and laydown is mixture-specific, but, in general, the current practice of loose-mix aging for 2 to 4 hours is adequate to simulate the changes that are observed in the field.

For the intermediate-temperature test, it was found that the variability in the within-lab and between-lab results at intermediate temperature (FI) continues to be a problem. While sample preparation was a challenge and might have contributed to some of the observed variability, the actual source of the high variability remains unknown. It was found that at least 8 samples should be tested to obtain an average that represents the actual value within 20%. This requires compaction of 2 gyratory pucks.

Notwithstanding the large coefficient of variation in the data, the Flexibility Index parameter can separate the extreme expected performers. Asphalt mixtures sampled at the plant

can be expected to have an FI generally between 3.0 and 20.0. The material sampled at laydown can be expected to have an FI somewhere between 2.1 and 18.5. Based on the literature, an FI limit between 6 and 10 would separate mixtures based on their expected performance. Application of this limit would result in 3 out of the 7 mixtures being eliminated. A small improvement in the differences between testing labs was seen when testing was done at a loading rate of 15 mm/min, but there was no clear benefit regarding the coefficient of variation. Furthermore, different loading rates resulted in different performance ranking of the mixtures. Unfortunately, without any performance information available from the field, it is not known if one loading rate is preferred over the other.

1.0 INTRODUCTION

1.1 Background

Utah DOT's pavements are its largest and most expensive asset. Within its current practice, UDOT is using aggressive rutting and stripping testing to qualify asphalt mixes for use in highway construction. This practice was in response to the typical distresses found in pavements from the late 1980s and early 1990s. In Utah, as well as in other states, this practice has generally resolved rutting issues, but has led to a detrimental effect on cracking and raveling behavior in the pavements. This one-dimensional approach has been recognized as a challenge to be addressed within the mix design process. Furthermore, in an attempt to resist rutting and be more environmentally friendly, mixes now contain Recycled Asphalt Pavement (RAP) and less asphalt binder, both virgin and total. In addition, with the high cost of asphalt binder and the increase in available substitutes and modifiers, mix performance testing is becoming increasingly important as binder testing alone is no longer adequate to predict mix performance.

Building a mix to avoid both rutting and cracking requires a balance of priorities since these behaviors are in direct conflict. However, in the absence of practical tests, the Utah DOT mix acceptance program currently favors rutting resistance, leaving a clear imbalance and skewed performance. As the program continues, these effects are becoming more pronounced. Should current practices continue without adjustment for durability performance, constructed pavements will continue to exhibit early age cracking (both thermal and fatigue) and the performance of the pavements will be significantly affected, leading to a significant loss of investment by the Department.

Two tests that evaluate low- and intermediate-temperature performance of asphalt mixtures were developed as part of previous research work. These tests use the existing Bending Beam Rheometer (BBR) to evaluate mixtures for low-temperature properties (modulus and relaxation capacity) and the Semi-Circular Bend (SCB) to evaluate mixtures for intermediate-temperature properties (Flexibility Index). By using these two tests (BBR and SCB), asphalt mixtures can be evaluated for cracking potential in regards to RAP content, binder content, binder modification, etc. resulting in a complete performance related specification that would exclude poor performing mixtures and allow for innovation in asphalt mix design.

1.1.1 Problem Statement

Before the adoption of any pavement performance related specification, it is important to understand ALL aspects of mixture design. This work is aimed at developing and understanding the durability properties (through low-temperature modulus, relaxation capacity and Flexibility Index) of mixes that are being produced in the state of Utah so that a limit value/specification can be set to ensure the best quality leading to longer lasting pavements. It is important to know what is being produced in the state in terms of low- and intermediate-temperature performance so that proper decisions can be made.

With the eventual establishment of design and field thresholds, these tests can be used to evaluate and classify mixes for cracking potential in regards to RAP content, binder content, binder modification, etc. thus allowing for innovation, better optimization of mixes, and reduction of poor performing pavements. The ranges and sensitivities established in this work are the first step in the development of thresholds to be included in the UDOT Standard Specifications.

1.2 Objectives

The objective of this research is to find appropriate tests limits/threshold values, based on actual mixture production, for the low- and intermediate-temperature properties of asphalt mixtures (i.e., low-temperature modulus, relaxation capacity, and Flexibility Index). Specific objectives are:

1. Determine the low-temperature properties of asphalt mixtures currently being produced in the state of Utah using the BBR test protocols to determine the range in the low-temperature modulus and the relaxation capacity of existing mixes.
2. Determine the intermediate-temperature properties of asphalt mixes currently being produced in the state of Utah by using the SCB test protocols to determine the range in Flexibility Index.

3. Determine how changes in mixture characteristics and aging between production and delivery to the paver affect the test parameters.
4. Suggest realistic acceptance limits for cracking that can be satisfied during both design and delivery of asphalt mixtures.

1.3 Scope

This study consists of the evaluation of asphalt mixture properties at two different temperature ranges (low and intermediate) using two different tests. One test, the BBR, addresses the cold temperature properties while the other, the SCB, addresses the intermediate-temperature properties of asphalt mixtures. Data was produced by preparing samples appropriate for each method and testing them based on established protocols or controlled testing variations.

The materials used in this study were obtained from asphalt mixture plants situated across the state of Utah intended for actual road construction. In the case of laboratory prepared materials, the mixtures used in previous research were utilized. While every effort was made to use consistent materials from year to year of this comprehensive study, it was recognized that the asphalt binder utilized as part of the laboratory study would be a standard material produced during the year in which testing was done and might have a slightly different composition from the one used in previous reports.

1.4 Outline of Report

This report is a continuation of the work previously described in the following research reports:

- *Development of Methods to Control Cold Temperature and Fatigue Cracking for Asphalt Mixtures* (Report No. UT-10.08) by Romero et al. (2011);

- *Using the Bending Beam Rheometer for Low Temperature Testing of Asphalt Mixtures* (Report No. UT-16.09) by Romero (2016);

- *Intermediate Temperature Cracking in HMA: Phase I Semi-Circular Bending (SCB) Practicality Evaluation* (Report No. UT-17.01) by VanFrank, et al. (2017); and
- *Balanced Asphalt Concrete Mix Performance Phase II: Analysis of BBR and SCB-IFIT Tests* (Report UT-17.21) by Romero and VanFrank (2017).

While some information is repeated in this report for clarity and ease of reading, most of the theoretical background has been omitted as it has already been presented in those reports. Readers are encouraged to read the previous reports available at the Utah Department of Transportation website: (www.udot.utah.gov/go/research).

This report is divided into the following sections:

- Introduction
- Literature Review
- Material Collection and Sample Preparation
- Evaluation of Field Produced Material Using the BBR Test
- Evaluation of Flexibility Index Using Laboratory Material
- Evaluation of Flexibility Index Using Field Material
- Summary, Conclusions, and Recommendations

2.0 LITERATURE REVIEW

2.1 Overview

Asphalt mixture is a complex material that, once placed in the field, is meant to withstand severe temperature extremes. In order to develop a more mechanistic-based approach to evaluate potential asphalt mixture performance, two test setups have been proposed, the Bending Beam Rheometer to evaluate low-temperature (thermal) cracking and the Semi-Circular Bend (SCB) to evaluate intermediate-temperature (fatigue) cracking. The development of these two tests is discussed here.

2.2 Low-Temperature Testing and Evaluation

A comprehensive literature review has been conducted on using the Bending Beam Rheometer to evaluate potential low-temperature performance of asphalt mixtures. Such information will not be repeated in this report but can be found in previous UDOT reports: UT-10.08, UT-16.09, and UT-17.21. As mentioned in Section 1.4, those reports are available at the Utah Department of Transportation website: (www.udot.utah.gov/go/research). Only a brief summary is provided here.

As a simple, fast, relatively inexpensive, and repeatable method of testing low-temperature properties of asphalt mixtures, researchers proposed the use of small beam specimens (12.7 mm width x 6.35 mm thickness x 127 mm length) made from asphalt concrete and tested on the Bending Beam Rheometer (BBR) [see References 1, 2]. It was shown that this testing configuration could be used to evaluate low-temperature properties in a way comparable to other mixtures tests such as the Indirect Tensile Test (IDT). These tests do not necessarily result in the same numerical value for creep modulus, but the comparison between the two of them is highly correlated [3]. Moreover, it has been demonstrated that both ‘creep modulus or stiffness’ and ‘stress relaxation capacity or m-value’ (slope of the logarithm of modulus vs. the logarithm of time curve) play a significant role in low-temperature performance of asphalt pavements. Asphalt concrete mixtures with high creep moduli and low m-values at their environmental design temperature (i.e., performance grade) are more susceptible to low-temperature thermal distress [4]. Success in these studies led to the formation of a provisional test standard for determining low-temperature properties of asphalt mixtures i.e., AASHTO TP125-16.

2.3 Intermediate-Temperature Testing and Evaluation

More information regarding intermediate-temperature testing can be found in previous UDOT reports UT-17.01 and UT-17.21 as mentioned in Section 1.4. A short summary is provided here.

In order to evaluate intermediate-temperature performance (i.e., fatigue cracking) in asphalt materials, many tests have been developed including indirect tension test (IDT), dissipated creep strain energy test (DCSE), four-point beam fatigue test (FBT), single-edge notched beam (SEB), disk-shaped compact tension (DCT), Texas overlay tester (OT) and semi-circular bending test (SCB).

Most of the fracture energy tests used to rank fracture toughness were developed by researchers in the field of rock or ice mechanics. Most of these tests are specified for cored based specimens with modifications to the Chevron bend specimen and short rod specimen [5]. In this manner, the SCB test was originally developed to determine the crack resistance and crack growth rate in rocks.

During the 1990s, the SCB test was proposed for bituminous mixtures. It was believed that this configuration was easier in comparison to other methods that were expensive and complex for regular use [6]. The SCB test gained some popularity for property characterization such as crack resistance by determining fracture toughness of asphalt mixtures in the early 2000's. The popularity of the test is due to simplicity in terms of specimen preparation using the Superpave Gyratory Compactor (SGC) or coring from the field [7, 8, 9]. Many researchers used SCB to study fracture properties of asphalt specimens at low temperature to differentiate cracking resistance [8, 9, 10, 11]. Standard protocols to unify different methods of SCB test at low temperatures such as EN12697-44: 2010 [12] and AASHTO TP105-2013 [13] were established. Recently, many researchers have studied the intermediate-temperature fracture resistance of various asphalt mixtures using the same SCB test configuration [14 – 21] leading to the development of parameters such as the Flexibility Index (FI).

The goal of the SCB development was to eliminate the mixtures that have a tendency for premature failure through a cracking related mechanism. In this manner, the asphalt mixtures would be tested in the laboratory prior to production. The ones characterized as improper by the fracture resistance properties would be eliminated. While this is an extension of the low-temperature testing, it is believed that the fracture resistance at intermediate temperature would result in better fatigue performance of the pavement in the field. Standard protocols have been developed for different methods such as ASTM D8044-2016 [22], known as the Louisiana method of SCB test, and AASHTO TP124-2016 [23], known as the Illinois Flexibility Index (I-FIT). These standards specify test procedures such as loading rate, specimen geometry and support conditions to obtain a value for fracture resistance.

Geometry and the loading configuration of SCB test based on the standards ensure that the tensile fracture (mode I) is dominant. Energy dissipation in SCB test is primarily governed by fracture mechanisms of crack initiation and crack propagation. In order to investigate the fracture mechanism in SCB, Arabani and Ferdowsi studied SCB tests and compared them to a suite of conventional tests such as indirect tensile strength test (ITS). It was observed that the SCB specimens fail with less distortion and a clear and anticipated crack path while the ITS test exhibits multiple modes of failure including wedging [14].

Promising configuration of SCB test and convenient fabrication of specimen from gyratory pucks encouraged Mull et al. to investigate the applicability of the SCB specimens by J_c characterization. Traditionally, J-integral values have been implemented by researchers as a tool to investigate fatigue crack growth of different materials. In a fatigue crack propagation study on asphalt mixture SCB specimens, Mull et al. investigated the energy release rate J_c from the fatigue hysteresis loops as a comparative tool. It was observed that compliance of the specimen increases with increased crack length [16].

Mohammad et al. investigated the sensitivity of J-integral values with varied notch depths and different asphalt mixtures to the indirect tensile stress and strain test results. Their study asserted that the concept of toughness (fracture toughness) is directly related to intermediate-temperature crack performance (fracture resistance) in pavements. The Louisiana State University Model (LSU) assumes that the energy (toughness) to move a crack at any point along the developing crack path is the energy under the stress-strain curve to the point that crack propagation commences. They observed that J-integral values from the semi-circular fracture test were sensitive to the change in asphalt binder type. It was found that the SCB measured Jc values demonstrated a good correlation with field cracking performance data [18, 19].

Al-Qadi et al. developed another test method that has been implemented to calculate fracture energy. This test is known as the Illinois Flexibility Index Test (I-FIT). They found that the results have consistent and repeatable trends corresponding to changes in asphalt concrete (AC) mixture design properties [20].

Nsengiyumva et al. investigated an experimental-statistical approach on SCB testing variables (i.e., the minimum recommended number of specimens, thickness, notch length, loading rate, and testing temperature) to evaluate fracture behavior of AC mixtures at intermediate service temperature conditions. Based on the test-analysis outcomes, it was concluded that the temperature of 21˚C, the loading rates of 0.1 to 0.5 mm/min, 5 mm length of the notch, thicknesses of 40 to 60 mm and a minimum of five to six samples are the statistical minimum to sufficiently represent fracture behavior of asphalt samples [21].

Many state departments of transportation (DOTs), including Utah DOT, have been trying to address balanced asphalt mixtures in order to reduce premature failures and to improve pavement performance by implementing the SCB standards. In this manner, VanFrank et al. evaluated the SCB test based on the Louisiana protocols and concluded that the standards provided trends that were consistent with the expected behavior. However, the sample preparation, especially the different notch lengths, and the data analysis based on ASTM D8044-2016 were deemed too difficult for routine testing. Romero and VanFrank evaluated more samples to determine if FI as proposed by Illinois could detect changes in mixture components. This study explored the effects of increased or reduced binder content, increased RAP content, and increased laboratory aging on the same materials [2]. The conclusions were that the FI can differentiate between different mixtures composition such as binder content, RAP content, and aging. Based on this work, FI was selected as a viable candidate to evaluate the intermediate-temperature performance of asphalt mixtures in Utah.

2.4 Summary

Based on the literature review, it is evident that the BBR and SCB can be successfully used to evaluate the cracking performance of asphalt mixtures. The parameters obtained from BBR testing, namely creep modulus and relaxation capacity (m-value) have been shown to relate to field performance. High modulus and low m-value would result in poor performing mixtures. The parameter obtained from SCB testing, namely Flexibility Index, has been shown to follow the expected trends that might result in poor performance. Both tests combined elements of mechanics-based analysis with some practicality to allow for adoption as routine tests.

3.0 MATERIAL COLLECTION AND SAMPLE PREPARATION

3.1 Overview

The objective of this work was to obtain a representative sampling of the asphalt concrete material that is produced in the state Utah and then measure their low- and intermediate-temperature properties using the proposed tests. Knowledge of the range of properties, and eventually the performance of the materials once placed in the field, will allow for the development of a specification limit capable of reducing the risk of early failure from cracking.

3.2 Material Selection

The state of Utah has a diverse climate and geology. As such it is expected that asphalt mixtures with different properties are produced across the state. Therefore, the plan for material selection consisted in identifying projects across the state where mixtures were being placed during spring and summer 2018 and where access was available within the available resources. In an effort to understand how different asphalt mixtures respond to the proposed testing, the range of mixtures was not limited to mixtures used by UDOT and, as such, mixtures with higher RAP content that normally would not be placed on UDOT roads were collected. Also, mixtures designed using both Marshall and Superpave methods were selected.

3.3 Materials Collection Process

Once the locations of materials production were identified, staff was sent to collect enough samples for testing. Material was collected at two locations: at the plant and at the field at laydown. At the plant, material was obtained from the conveyor slat as it came from the mixer; while at laydown, the material was collected from the windrow dump. For all cases, the material was placed in 5-gallon metal buckets and sealed while still hot. The temperature of the material at sampling was recorded. The material was then transported to a central location where it was distributed to the three testing labs.

3.4 Material Properties

The material collected as part of this study had the following properties based on the mixture design information provided by the producers (see Table 3-1). The name of the producers, or the exact location where the mixtures were obtained and placed is not provided in this report in an effort to respect proprietary business information.

Table 3-1 Description of Field Materials Collected

Mix ID	Design Method	Aggregate NMAS	RAP Content	Total Binder by Mass	Virgin Binder by Mass/ Vol	Virgin Binder	Intended Climate
UT-01	50-Blow Marshall[1]	12.5 mm	30%	5.4%	3.8%/ 9.0%	PG 64-22	Hot
UT-02	75-Blow Marshall[1]	19 mm	30%	4.9%	3.4%/ 9.6%	PG 58-34	Medium
UT-03	75-NDES Superpave[2]	12.5 mm	25%	5.3%	4.0%/ 9.6%	PG 64-34	Cold
UT-04	75-NDES Superpave[2]	12.5 mm	15%	5.3%	4.6%/ 10.9%	PG 64-34	Medium
UT-05	50-Blow Marshall[1]	12.5 mm	30%	6.3%	4.4%/ 10.1%	PG 58-28	Cold
UT-06	75-NDES Superpave[2]	12.5 mm	25%	4.8%	3.7%/ 11.2%	PG 58-28	Cold
UT-07	75 NDES Superpave[2]	12.5 mm	10%	5.3%	4.9%/ 11.1%	PG 64-28	Medium

1. Based on APWA specifications
2. Based on UDOT 2741 specification

Note: all information provided by supplier and not verified by research team

3.5 Sample Preparation

As described in Section 3.3, the material was collected and brought to a central location. Buckets were distributed to 3 different laboratories: University of Utah, PEPG, and UDOT Central Materials Lab. These labs will be referred to in this report as Lab A, Lab B, and Lab C, respectively.

Given that the amount of materials was limited, Lab B performed extensive volumetric testing on each mixture to determine the maximum theoretical specific gravity, G_{mm}, of each mixture. Knowing the G_{mm} was necessary so that the right amount of material could be added to the Superpave gyratory compactor to achieve the target air voids. However, even with all the volumetric testing at one lab, the process still required a trial-and-error process until the right quantities were determined. Small variation in material made it sometimes difficult for the other labs to achieve the exact target air voids during compaction of their first gyratory cylinder; most materials required a second compaction. Lab A utilized the first compacted cylinder to evaluate the effect of air voids and other 'out of spec' variables; Lab B only reported testing results from the final compaction; while Lab C only reported one specimen. Once ready, the samples were cut for SCB testing in each of the three labs. All three labs found that producing consistent cuts was difficult and many of the samples did not meet the production standards defined by the test procedure. These inconsistencies may have introduced some level of variability to the test values; however, all of the results were used when analyzing results. Lab A also tested the material in the BBR for low-temperature performance; Lab C did limited BBR testing to verify test results.

4.0 EVALUATION OF FIELD PRODUCED MATERIAL USING THE BBR TEST

4.1 Overview

As described in Section 3, asphalt mixtures from seven different projects were investigated. For each project, the materials were gathered at two locations, at the plant and at the field at lay-down. Due to transportation from the plant to the site, lay-down material has short-term aged when compared to the plant material. The time between plant and laydown was normally less than 1.5 hours with one exception so no large variations in aging time exist. However, it has been shown that the effect of aging is actually mixture dependent so different properties are expected for all mixtures. The details of the seven mixtures were presented in Table 3-1. All BBR testing was done at Lab A with limited testing done at Lab C.

4.2 Sample Preparation and Testing

Once the materials were received in the lab, the collected mixture was cataloged and the material available was weighed. Based on the information provided by the producer and the volumetric results measured at Lab B, enough material was weighed so that a 110-mm high gyratory cylinder could be compacted to 4 ± 1 % air voids. This range of air voids was used since previous studies have indicated that the results from the BBR are not particularly sensitive to air voids within that range. For each mixture a gyratory cylinder was compacted and allowed to cool. Then it was cut into small beams based on the protocols described in AASHTO TP125-16. Over twenty samples were obtained from each cylinder out of which the best 12 in terms of consistent dimensions were selected for testing.

The small beams were conditioned following the protocols described in previous studies [1, 2]. Each beam was tested at 3 temperatures in order of coldest to warmest. Previous studies have demonstrated that repeated testing of the same beam is acceptable and does not affect the results [24]. The test temperatures were -24 ºC, -18 ºC, and -12 ºC to represent the low-temperature performance grade environments seen in Utah (PG XX-34, PG XX-28, and PG XX-22) and not necessarily the binder performance grade used in the mix. This was done so that the mix could be evaluated for a given environment and in recognition to the fact that the 'true' binder grade of the mix once binder, aging, aggregates, and RAP interact is not known.

4.3 Results

Each mixture was tested in creep load using the BBR at each temperature for 180 seconds. This provides the complete time- and temperature-dependent creep modulus of the material. However, given that the specification only requires data at 60 seconds, the data was summarized for that specific time. For each test, the creep modulus (referred to as modulus in this document for simplicity) and the relaxation capacity, or m-value (slope of the log-modulus log-time curve at the given time), were determined. The values for the twelve samples tested for each mix were averaged and the standard deviation was determined. These results are presented in Table 4-1 and Table 4-2.

As can be seen in the tables, the results are fairly consistent with a coefficient of variation below 25% in all cases and, in most, even below 15%. The data shows the expected trend of decreasing modulus and increasing m-value as the temperature increases. In four out of the seven sections (UT-02, UT-03, UT-06, and UT-07), there is an increase in modulus and a decrease in m-value between the material collected at plant and the material collected in the field indicating that short-term aging occurred. In two sections (UT-01 and UT-05) there is not a clear indication of aging as the results are within the margin of error. Of concern are the results for UT-04 which shows an unexpected trend of decrease in modulus and increase in m-value between plant and field collection. Since this was unexpected, tests were run in Lab C at -24 ºC. The results from Lab C also show a decrease in modulus (11,569 MPa versus 10,520 MPa) and an increase in m-value (0.118 versus 0.141) for the same sampling locations thus confirming the results of the test. It is unclear if these results are indeed a representation of some actual physical behavior, some outlying result, or a labeling mistake.

Table 4-1 BBR Results for Projects 01 through 04

	Testing Temperature, °C	-24		-18		-12	
	Sampling Location	**Plant**	**Field**	**Plant**	**Field**	**Plant**	**Field**
	Sample Size (n)	12	12	12	12	12	12
UT-01	**Average Modulus at 60s (MPa)**	**18,192**	**17,362**	**14,442**	**14,583**	**11,460**	**11,505**
	Standard Deviation, s	2089.57	4070.27	1532.95	1673.77	1353.08	1862.70
	Coefficient of Variation	0.11	0.23	0.11	0.11	0.12	0.16
	Average m-value at 60s	**0.089**	**0.098**	**0.123**	**0.110**	**0.166**	**0.147**
	Standard Deviation, s	0.0080	0.0165	0.0080	0.0125	0.0087	0.0141
	Coefficient of Variation	0.09	0.17	0.07	0.11	0.05	0.10
UT-02	**Average Modulus at 60s (MPa)**	**16,692**	**17,808**	**14,075**	**14,958**	**10,562**	**11,437**
	Standard Deviation, s	2042.93	3766.71	2230.42	2731.11	1213.21	1819.95
	Coefficient of Variation	0.12	0.21	0.16	0.18	0.11	0.16
	Average m-value at 60s	**0.087**	**0.106**	**0.118**	**0.118**	**0.158**	**0.152**
	Standard Deviation, s	0.0088	0.0143	0.0130	0.0121	0.0097	0.0138
	Coefficient of Variation	0.10	0.13	0.11	0.10	0.06	0.09
UT-03	**Average Modulus at 60s (MPa)**	**14,033**	**15,133**	**9,339**	**9,743**	**6,253**	**6,648**
	Standard Deviation, s	1867.59	1181.94	1431.96	2546.22	1297.29	1669.68
	Coefficient of Variation	0.13	0.08	0.15	0.26	0.21	0.25
	Average m-value at 60s	**0.126**	**0.121**	**0.170**	**0.169**	**0.241**	**0.242**
	Standard Deviation, s	0.0127	0.0128	0.0102	0.0163	0.0220	0.0155
	Coefficient of Variation	0.10	0.11	0.06	0.10	0.09	0.06
UT-04	**Average Modulus at 60s (MPa)**	**13,308**	**10,715**	**10,228**	**7,855**	**6,264**	**4,189**
	Standard Deviation, s	1302.07	2385.37	764.09	1437.04	763.13	725.95
	Coefficient of Variation	0.10	0.22	0.07	0.18	0.12	0.17
	Average m-value at 60s	**0.130**	**0.162**	**0.188**	**0.220**	**0.259**	**0.298**
	Standard Deviation, s	0.0130	0.0188	0.0080	0.0233	0.0106	0.0328
	Coefficient of Variation	0.10	0.12	0.04	0.11	0.04	0.11

Table 4-2 BBR Results for Locations 05 through 07

	Testing Temperature, °C	-24		-18		-12	
	Sampling Location	**Plant**	**Field**	**Plant**	**Field**	**Plant**	**Field**
	Sample Size (n)	12	12	12	12	12	12
UT-05	**Average Modulus at 60s (MPa)**	**20,083**	**19,917**	**17,167**	**15,408**	**12,408**	**11,921**
	Standard Deviation, s	2251.80	1444.63	3235.41	1981.02	1171.21	1580.20
	Coefficient of Variation	0.11	0.07	0.19	0.13	0.09	0.13
	Average m-value at 60s	**0.100**	**0.099**	**0.125**	**0.126**	**0.166**	**0.178**
	Standard Deviation, s	0.0130	0.0086	0.0141	0.0150	0.0128	0.0114
	Coefficient of Variation	0.13	0.09	0.11	0.12	0.08	0.06
UT-06	**Average Modulus at 60s (MPa)**	**14,507**	**20,808**	**12,101**	**16,225**	**8,043**	**13,125**
	Standard Deviation, s	3614.04	3749.05	2608.50	2414.02	1789.09	2012.74
	Coefficient of Variation	0.25	0.18	0.22	0.15	0.22	0.15
	Average m-value at 60s	**0.105**	**0.094**	**0.141**	**0.114**	**0.178**	**0.161**
	Standard Deviation, s	0.0113	0.0120	0.0129	0.0080	0.0189	0.0120
	Coefficient of Variation	0.11	0.13	0.09	0.07	0.11	0.07
UT-07	**Average Modulus at 60s (MPa)**	**12,479**	**14,683**	**9,836**	**11,686**	**6,061**	**7,335**
	Standard Deviation, s	3188.65	1888.64	2285.15	2284.05	1159.33	1405.58
	Coefficient of Variation	0.26	0.13	0.23	0.20	0.19	0.19
	Average m-value at 60s	**0.138**	**0.122**	**0.189**	**0.167**	**0.248**	**0.243**
	Standard Deviation, s	0.0170	0.0119	0.0260	0.0140	0.0281	0.0200
	Coefficient of Variation	0.12	0.10	0.14	0.08	0.11	0.08

Previous research [4] has recommended values below 12,000 MPa for modulus and m-values above 0.12 as a criterion for low-temperature performance. Based on the modulus limit alone, none of the mixtures tested is appropriate for a PG XX-34 environment, two (UT-03 and UT-07) are appropriate for a PG XX-28 environment and just about all of them are adequate for a PG XX-22 environment. Using only the m-value as a criterion, UT-03, UT-04, and UT-07 might be appropriate for the PG XX-34 environment, and all, except UT-02, might be appropriate for the PG XX-28 environment; all mixtures have m-values above 0.12 at PG XX-22. However, performance is not based on one single value; research has actually shown that high modulus might have acceptable performance as long as the m-value is high.

To further understand the behavior of the mixtures, Black Space diagrams (modulus in the y-axis and m-value in the x-axis) were used to graphically show the changes in modulus and m-value with temperature and with short-term aging. See Figure 4-1 for an example. Points that fall inside the shaded area represent the values that would likely result in premature cracking. Note that the shaded area is triangular in shape indicating that high modulus is acceptable if the m-value is also high. Each project is described in detail in the next section; readers not interested in such details can skip to Section 4.6 for a summary of results.

4.3.1 UT-01

Figure 4-1 shows the results for a mix designed using the Marshall method and containing relatively high amounts of RAP (30%). The graph shows that this mix would not be expected to perform well in a PG XX-34 ºC or a PG XX-28 environment but should work in warmer environment (PG XX-22 ºC).

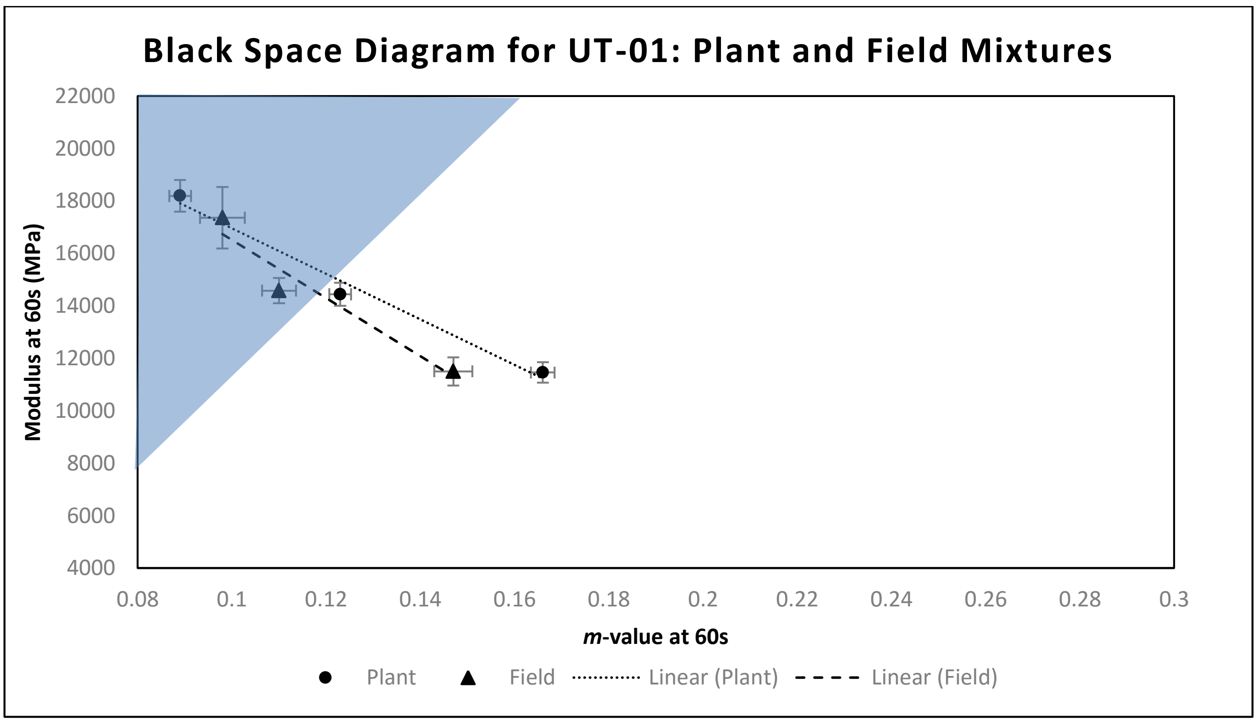


Figure 4-1 Low-Temperature parameters for UT-01

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

The data shows an increase of modulus and a decrease of *m*-value, as is expected, with aging of the material. However, the difference in modulus is not significant except at -28 °C, but the decrease in m-value at this temperature makes the mix not adequate for the PG XX-28 environment.

<u>4.3.2 UT-02</u>

Figure 4-2 shows a mixture similar to UT-01 in terms of design method and RAP content, albeit different Marshall blows. Like the previous mix, the results show that this material is not suitable for a PG XX-34 ºC environment. When comparing both mixtures, it is evident that the use of a softer virgin binder (PG 64-22 in UT-01 vs. PG 58-34 in UT-02) results in a 'softer' mixture at -24°C although not soft enough to be considered acceptable in a PG XX-34 environment. No significant difference between plant and field (lay down) mixtures is seen except at -24°C, where m-value of the plant mixture is lower than the field mixture and the creep modulus is higher.

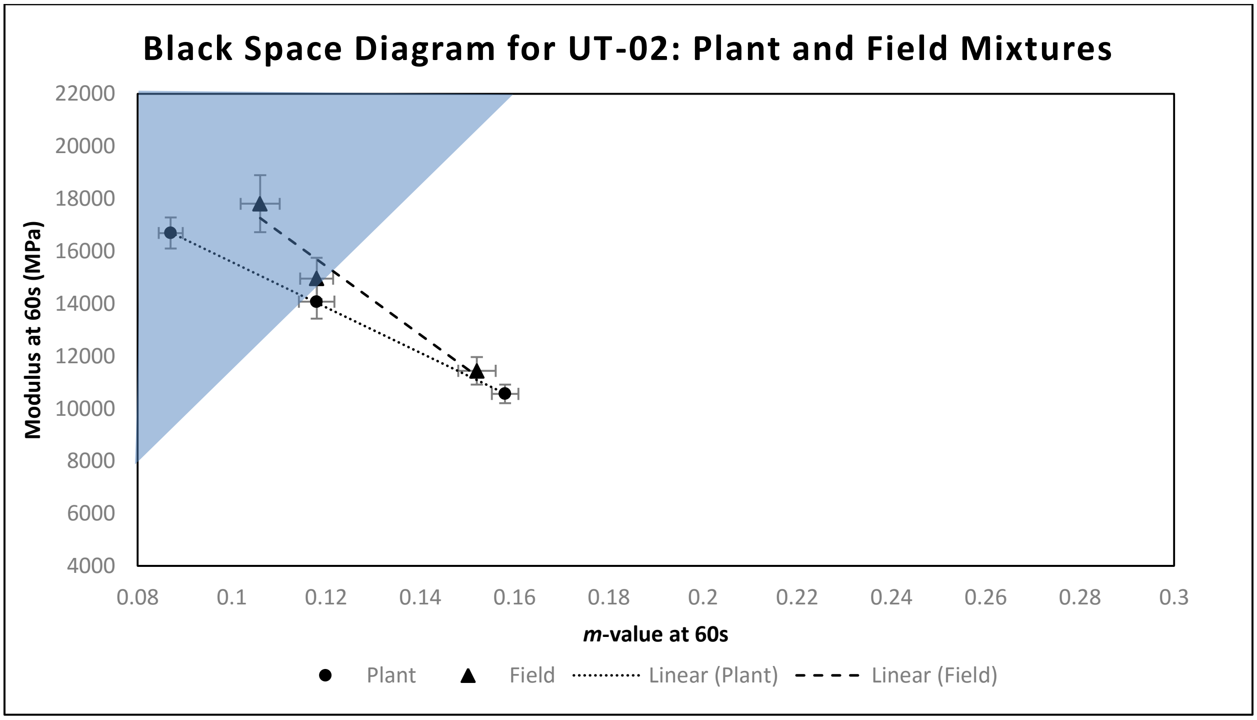


Figure 4-2 Low-Temperature Parameters for UT-02

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.3.3 UT-03

Figure 4-3 shows the results for material designed using the Superpave method with a RAP content of 25%. While the modulus is above the 12,000 MPa for a PG XX-34 environment, the m-value is close to the 0.12 limit making it possible that it might resist premature cracking. At higher temperatures, this mix would be expected to perform satisfactorily. It is believed that the main difference between this mix and the previous two, besides having 5% less RAP, is a higher binder content. For this mix, the differences between the results of plant and field samples are not significant.

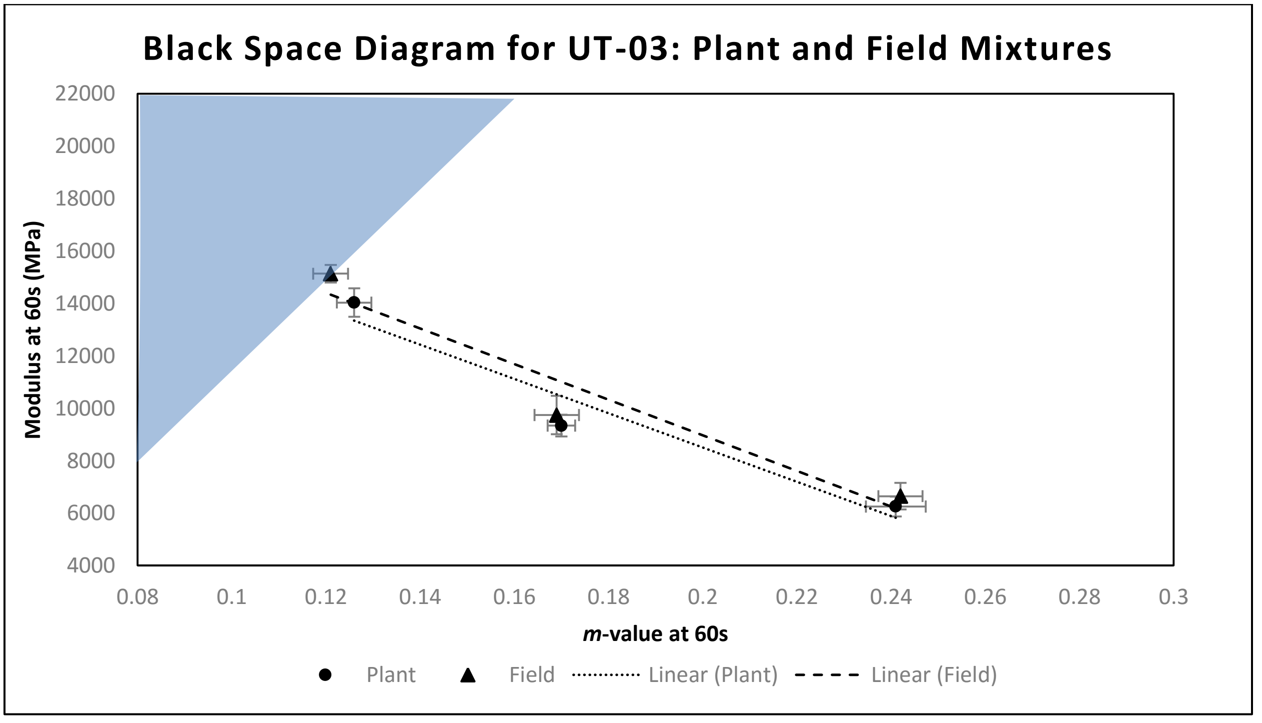


Figure 4-3 Low-Temperature Parameters for UT-03

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.3.4 UT-04

Figure 4-4 shows the graph for a mixture very similar to UT-03 but with only 15% RAP. While there is not a significant difference between UT-03 and UT-04 in term of modulus, the mixture with a lower RAP content also has higher m-value. In general, this mix shows low modulus and high m-value, making it appropriate for all environments evaluated. From a low-temperature performance perspective, this is one of the best mixtures of the seven projects evaluated. However, as previously discussed, the lay down mixture has lower creep modulus and higher *m*-value compared to the plant mixture. This is contrary to theoretical expectations. These results were verified by a second lab indicating that it is a real physical behavior; however, the reasons for this anomaly are not known.

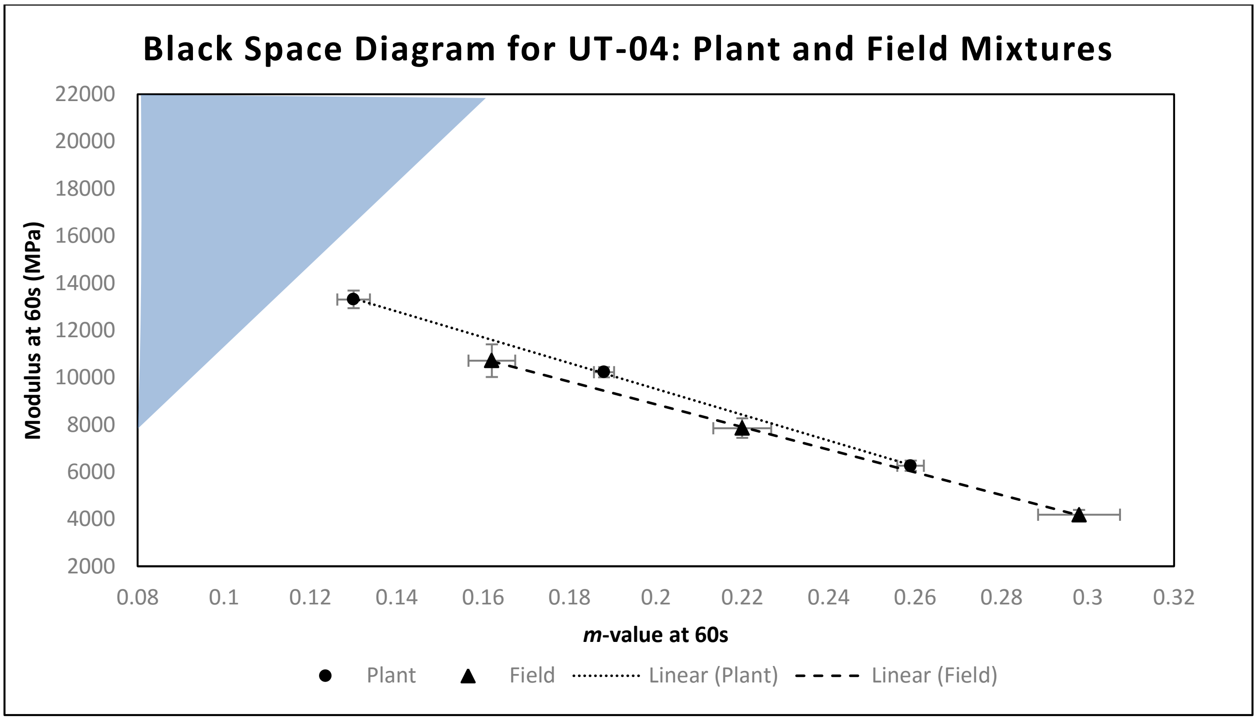


Figure 4-4 Low-Temperature Parameters for UT-04

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.3.5 UT-05

Figure 4-5 shows the results for a mixture design using the Marshall procedure with 30% RAP content. Based on the RAP content, its performance can be compared to UT-01 even though the virgin binder grade (PG 64-22 for UT-01 and PG 58-28 for UT-05) and content (3.8% virgin binder for UT-01 and 4.4% for UT-05) are different. However, based on the results, both mixtures would be expected to have very similar performance in the same environments and only be adequate in warm environments (PG XX-22). The test results are consistent with previous experience and the expectation that mixtures with 30% RAP could not adequately perform at low temperatures.

This project was the only one in which there was a significant time difference between the collection of the mixture at the plant and at the field. After mixing, the material was stored at the silo overnight making the time between plant and laydown over 17 hours. At -12 ºC there is no change in modulus but the m-value increases; at -18 ºC there is no change in m-value but the modulus decreases. A t-test run on the creep modulus at -18 ºC resulted in a P-value of 0.1225 indicating that the difference in results is considered to be not statistically significant at 95% confidence. Finally at -24 ºC, there is no difference between plant and field.

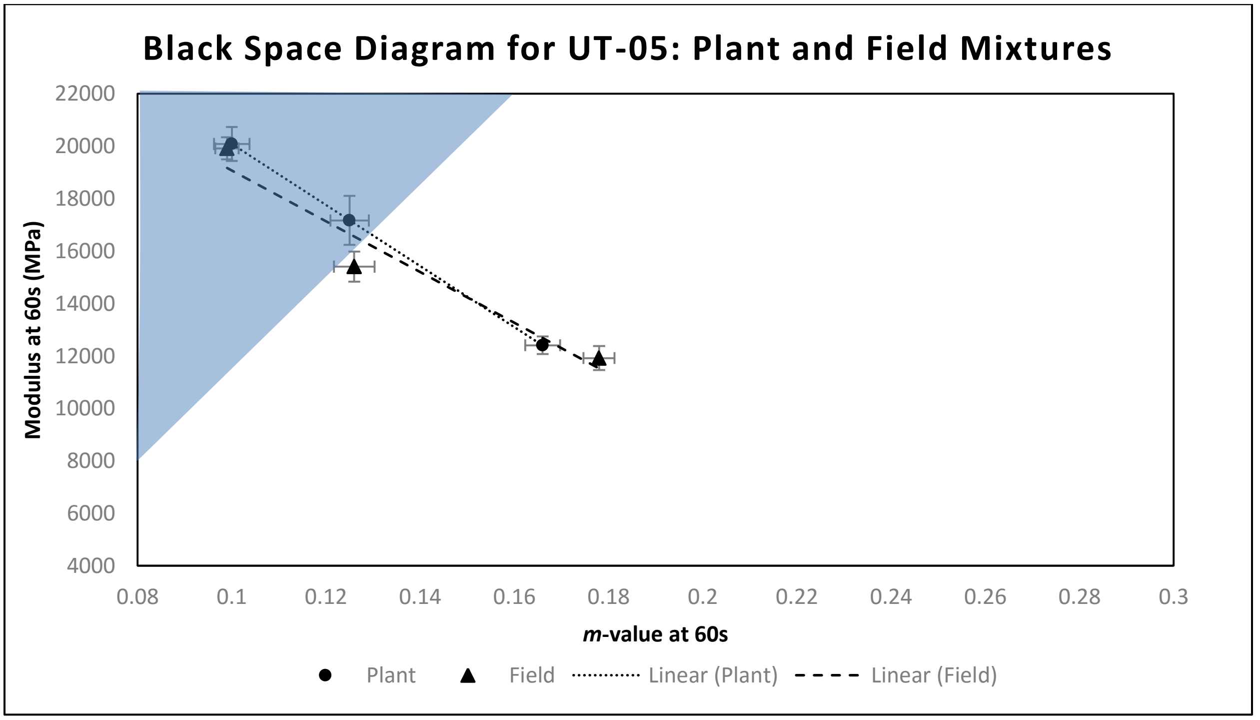


Figure 4-5 Low-Temperature Parameters for UT-05

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.3.6 UT-06

Figure 4-6 shows the results for UT-06 mixture. This mixture has the same 25% RAP content as UT-03 so comparisons between the two are of interest. UT-03 has lower creep modulus and higher m-value than this mix. This is expected since UT-03 uses a 'softer' low-temperature performance binder grade (PG 64-34 versus PG 58-28 for UT-06) and higher total binder content.

This mixture is very susceptible to aging; a large difference in results is observed in both the creep modulus and the m-value between the samples collected at the plant and the samples collected at laydown. At -18 ºC, this difference causes the aged samples to be considered inadequate for the PG XX-28 environment.

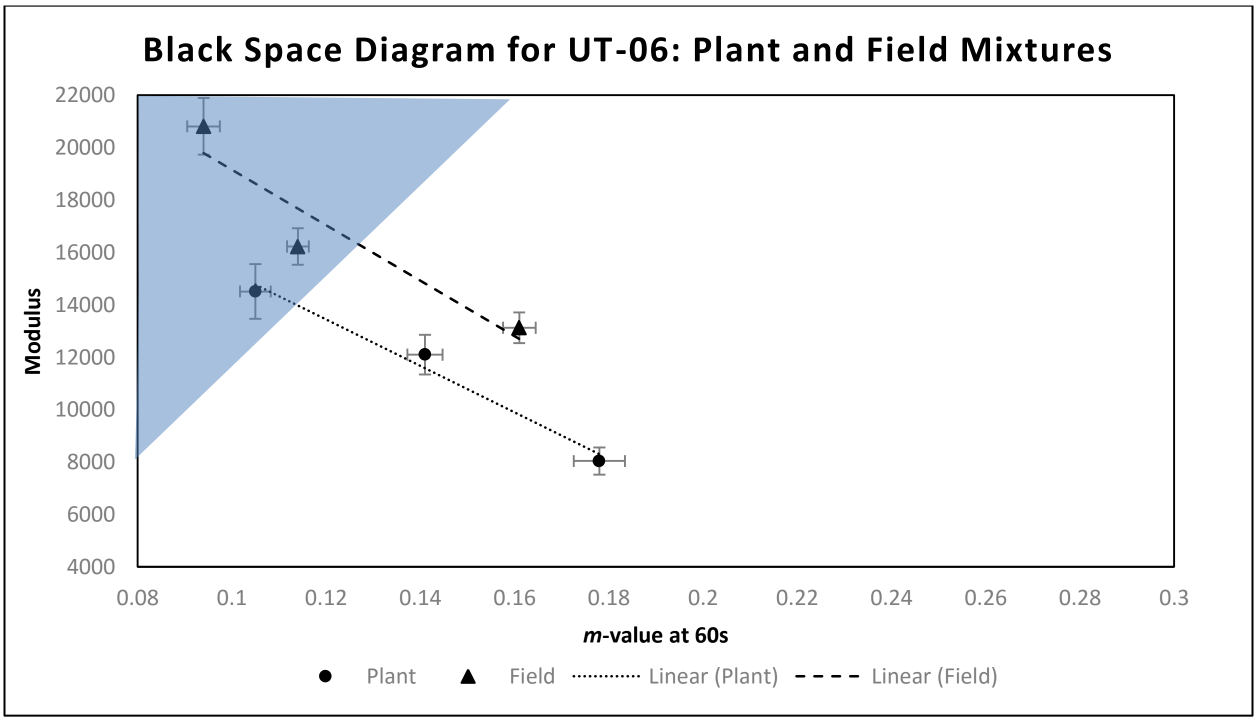


Figure 4-6 Low-Temperature Parameters for UT-06

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.3.7 UT-07

Figure 4-7 shows the results for UT-07 mixture. The asphalt mixture used in project UT-07 has the least amount of RAP (10%) and highest amount of virgin binder content (4.9%) among all mixtures. At -24 ºC and -18 ºC it shows significant aging; however, even after the short-term aging, this mixture would be expected to have satisfactory performance even at a PG XX-34 environment. No other mixture prepared with a PG 64-28 virgin binder has been found to be adequate for the lowest temperature. This demonstrates the value of mixture testing since binder testing alone could not have suggested satisfactory performance in the lowest temperature environment.

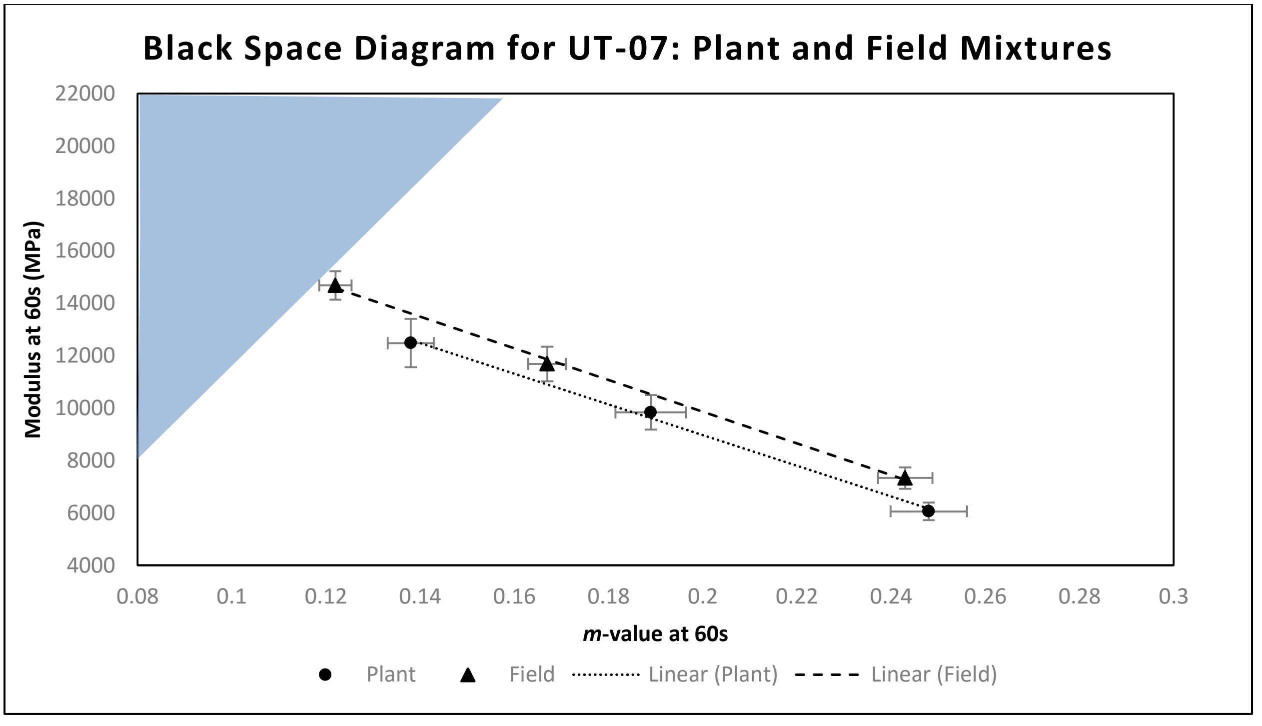


Figure 4-7 Low-Temperature Parameters for UT-07

Low-temperature parameters at -24°C, -18°C, and -12°C (left to right). Error bars represent standard error of the means of modulus and *m*-value. Shaded area represents values outside the proposed limits.

4.4 Effect of Short-Term Aging

The data collected as part of this study allows for the comparison of the effect of short-term aging between the plant and laydown. As was shown in the previous sections, in 5 out of the 7 sections, the short-term aging resulted in an increase in creep modulus and a decrease in the m-value. Previous research [2] has recommended using an aging index parameter to represent the combined changes in both creep modulus and m-value. The same research indicates that different mixtures will age differently based on aging time, temperature, and RAP content.

The aging index is calculated from the change in modulus, Δ_{modulus}, and the change in m-value $\Delta_{\text{m-value}}$. Calculations for this parameter are based on the following equations.

$$\Delta_{modulus} = \frac{Modulus_{Field} - Modulus_{Plant}}{Modulus_{Plant}} \quad \text{Equation 1}$$

$$\Delta_{m-value} = \frac{m-value_{Field} - m-value_{Plant}}{m-value_{Plant}} \quad \text{Equation 2}$$

$$Index = \sqrt{(\Delta_{modulus})^2 + (\Delta_{m-value})^2} \quad \text{Equation 3}$$

The results of the aging index for the tests done at -18°C are shown in Table 4-3. As previously discussed, sections UT-04 and UT-05 are omitted due to irregularities in the results. Mixtures from two projects (UT-06 and UT-07) show significant aging between plant and laydown.

Table 4-3 Aging Index at -18 °C

	Creep Modulus MPa		m-value		Δ_{modulus}	$\Delta_{\text{m-value}}$	Aging Index
	Plant	Field	Plant	Field			
UT-01	14,442	14,583	0.123	0.110	0.010	-0.106	11%
UT-02	14,075	14,958	0.118	0.118	0.063	0.000	6%
UT-03	9,339	9,743	0.170	0.169	0.043	-0.006	4%
UT-04*	10,228	7,855	0.188	0.220	-0.232	0.170	*
UT-05*	17,167	15,408	0.125	0.126	-0.102	0.008	*
UT-06	12,101	16,225	0.141	0.114	0.341	-0.191	39%
UT-07	9,836	11,686	0.189	0.167	0.188	-0.116	22%

*values where the creep modulus decreased or the m-value increased with aging.

4.4.1 Comparison of Field Aging and Laboratory Aging

The difference between the mixture properties before and after short-term aging can be compared to the laboratory aging that was done as part of previous research [2]. That work recommended extended aging of the loose mix at the compaction temperature. Depending on the RAP content, 1 hour of loose-mix aging will result in an increase in the Aging Index between 2 to 4 percent. This means that, based on the data collected in this study, less than 3 hours would simulate the aging that occurred in mixtures from sections UT-01, UT-02, and UT-03 but more are needed for mixtures for sections UT-06 and UT-07. The range in results is an indication that the effect of aging is mixture dependent. Furthermore, the trends observed are consistent with previous work that shows mixtures with high RAP content having a lower Aging Index than mixtures with low RAP content.

4.5 Multi-Lab Comparison

In order to evaluate the between-labs variability of the tests, limited testing on the same material was performed at Lab C. Previous comparisons between the two labs have shown that a difference of less than 10% would be expected [24]. The results, shown in Table 4-4, show a significant difference in the results obtained at the two labs. On average, a difference of 27% in the creep modulus was measured; yet on the same data, the average difference in the m-value was only 3% with only 3 out of 14 tests having an absolute difference greater than 15%. Given that, in the past, multiple comparisons have been done between Lab A and Lab C that resulted in values that were closer to each other, and that it is unlikely that the modulus would be different when the m-value is not, an investigation of the procedures followed by both laboratories was done. This consisted in staff from both labs visiting and discussing the procedures that were followed during specimen fabrication and testing. No specific cause was found but it is known that small changes in air voids do not have a significant effect in the results. It is also known that steric aging does not play a role after a few hours; therefore, an error in measurement was suspected.

Table 4-4 BBR Results at Lab A and Lab C

Section	Sampling Location	Test Temp, ºC	Creep Modulus, MPa Lab A	Lab C	Diff	m-value Lab A	Lab C	Diff
UT-01	Plant	-12	11,460	7,959	31%	0.166	0.149	10%
	Field		11,505	8,257	28%	0.147	0.108	26%
UT-02	Plant	-18	14,075	8,137	42%	0.118	0.128	-9%
	Field		14,958	8,358	44%	0.118	0.125	-6%
UT-03	Plant	-24	14,033	10,473	25%	0.126	0.120	5%
	Field		15,133	14,769	2%	0.121	0.149	-23%
UT-04	Plant	-24	13,308	11,569	13%	0.130	0.118	10%
	Field		10,715	10,520	2%	0.162	0.141	13%
UT-05	Plant	-18	17,167	9,705	43%	0.125	0.136	-8%
	Field		15,408	10,013	35%	0.126	0.115	9%
UT-06	Plant	-18	12,101	9,503	21%	0.141	0.136	4%
	Field		16,225	9,635	41%	0.114	0.122	-7%
UT-07	Plant	-18	9,836	6,371	35%	0.189	0.176	7%
	Field		11,686	9,398	20%	0.167	0.136	19%
				Average	27%		Average	3%
				Max	44%		Max	26%

The first observation from Table 4-4 is the fact that the difference does not appear to be random. There is a bias in the data where Lab C results are always lower by an average of approximately 4000 MPa when compared to Lab A. As previously mentioned, given that the m-value does not show such bias, an error in either load or deflection measurement is suspected. If a value of 4000 MPa is added to Lab C's data, then the difference in the modulus between both labs decreases to an average of 3% with only 3 out of 14 sections showing a difference greater than 17%.

A comparison of the results both before and after the correction was applied is shown in Figure 4-8. The corrected values show better agreement between the two labs.

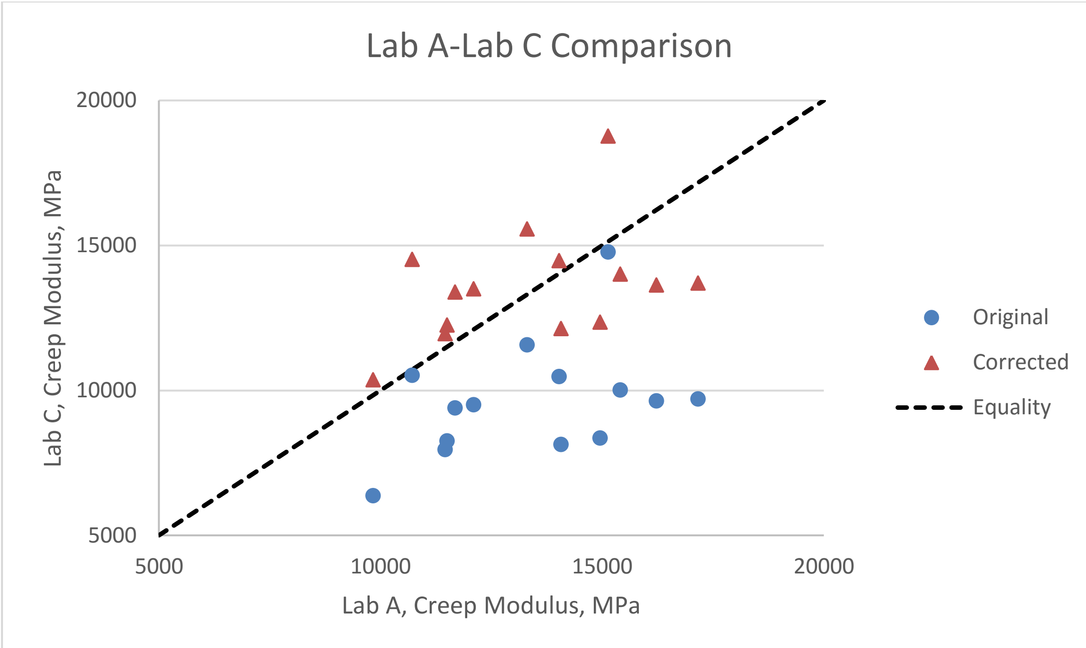


Figure 4-8 Comparison of Creep Modulus between Lab A and Lab C

In order to verify that testing in both labs provides the same results once any errors were corrected, ten new beams from a single mixture (different from the ones already evaluated) were tested at both labs. The results are shown on Table 4-5.

Table 4-5 Results from Lab A and Lab C on a New Material

Lab A		Lab C		Difference (A-C)	
Modulus MPa	m-value	Modulus MPa	m-value	Modulus	m-value
8,974	0.156	9,191	0.159	-2.4%	-1.7%

As shown on Table 4-5, a difference of less than 3% between both labs is observed for both the modulus and the m-value. This is consistent with previous work.

4.6 Summary

The results from low-temperature testing of asphalt mixtures collected from 7 field projects were shown in this section. While most mixtures have a modulus that might be considered too high for low-temperature cracking performance, the combination of modulus and m-value might allow them to adequately perform. The expected performance is presented in Table 4-6.

Table 4-6 Predicted Performance at Different Environments

	Low-Temperature Environment		
Mixture	**PG XX-34**	**PG XX-28**	**PG XX-22**
UT-01*	Fail	Fail	Pass
UT-02*	Fail	Pass	Pass
UT-03	Pass	Pass	Pass
UT-04	Pass	Pass	Pass
UT-05*	Fail	Fail	Pass
UT-06	Fail	Fail	Pass
UT-07	Pass	Pass	Pass

*Mixtures with 30 % RAP

Comparison of the mixtures properties after short-term aging indicates the test is capable of quantifying the changes observed in the field that are mixture-specific. In one case (UT-05), even after 17 hours of storage, the modulus and m-value did not change significantly; but, in another case (UT-06), 1.5 hours was enough to significantly change the properties between plant and laydown. This difference in the amount of aging measured for different mixtures is consistent with previous reports that showed certain combination of virgin binder and RAP to be more susceptible to aging.

4.6.1 BBR Conclusion

Based on the results from BBR testing, it is apparent that the low-temperature properties of the mixes are dependent on the interaction between the asphalt binder type, grade, and amount; the type and amount of RAP; the aging environment and time; and, possibly, the aggregate type and gradation. This, in itself, shows the value of the BBR for asphalt mixtures testing since it evaluates the properties when all components are combined thus allowing for a better performance expectation. The results also indicate that, as long as the equipment is properly calibrated and procedures closely followed, the results are repeatable between labs.

5.0 EVALUATION OF FLEXIBILITY INDEX USING LABORATORY MATERIAL

5.1 Introduction

In order to evaluate the variability of SCB-FI results, a controlled laboratory study was conducted prior to testing field material. The purpose of the study was to quantify the variability that would be expected as a results of materials testing only (i.e., not material preparation) and to ensure that all participating labs could produce the same results.

5.2 Methodology

As explained in Section 3.5, three different laboratories were used, each with their own locations, technicians, and equipment. In order to understand the variability of SCB-FI testing only, the effect of mixing, compaction, and sample preparation was held constant to isolate this potential source of variability. Ensuring consistency of sample preparation avoids confusing the laboratory variability with the material variability. Three pucks of a single gradation mixture with one source of binder and without RAP or any other additives were produced in Lab C by one operator. To be consistent with the thickness of cuts and notch of the SCB sample, the whole cutting process was done in Lab A by one operator.

The aggregates used were obtained from one source with a nominal maximum aggregate size of 12.5 mm. The binder grade was PG64-34 and the binder content was 5.3% by total mix weight. The mixture was designed at 75 gyrations as a compaction specification, but compaction was performed to height to achieve 7.0±0.5 % air void. Specific details of this mix are shown in a previous report as Mix A (Report No. UT-17.21). Three cylindrical SGC samples were brought to Lab A where they were cut into twelve SCB cut specimens. The twelve samples were randomized and four SCB samples were distributed to each of the three laboratories. All laboratories tested their four SCB samples with a similar instrument in terms of brand and functionality on the same day. The whole procedure was repeated on three consecutive days.

SCB samples were conditioned at 25˚ C (testing temperature) for 1hr ± 5min to achieve constant sample temperature during the test. An incubator was used in each laboratory. A dummy sample connected to a digital thermometer was paired with the test sample to ensure that conditioning temperature and the temperature of the sample during the test were consistent. The SCB machine used for this study is an Illinois – Flexibility Index Tester (I-FIT) and the testing was done in accordance with AASHTO TP124-16 with minor exceptions regarding the loading rate. The SCB machine fixture at Lab A, enclosed inside a chamber, is shown in Figure 5-1. The chamber helps to eliminate air currents to maintain a temperature of 25 ºC during the test. Three different thermometers including the dummy puck thermometer were used to assure that the temperature inside the chamber was consistent.

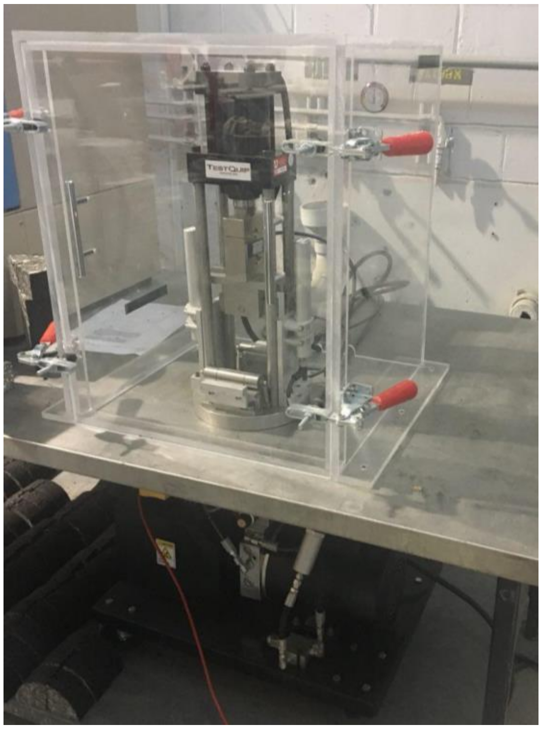


Figure 5-1 SCB Instrument enclosed in a chamber at Lab A.

Flexibility Index (FI) values were calculated to evaluate the repeatability (day-to-day) and the reproducibility (lab-to-lab) of the SCB-FI test. The FI calculations were done automatically using the provided machine software for all testing. In all cases, the FI values were calculated using Equation 4.

$$FI = A \times \frac{G_f}{|m|} \qquad \text{Equation 4}$$

In this equation G_f is work of fracture obtained as the area under the load-displacement curve divided by the ligament area. The value $|m|$ is the absolute value of the slope of the post-peak curve at the inflection point reported as kN/mm. Coefficient A is a calibration coefficient for unit conversions and possibly field aging shift.

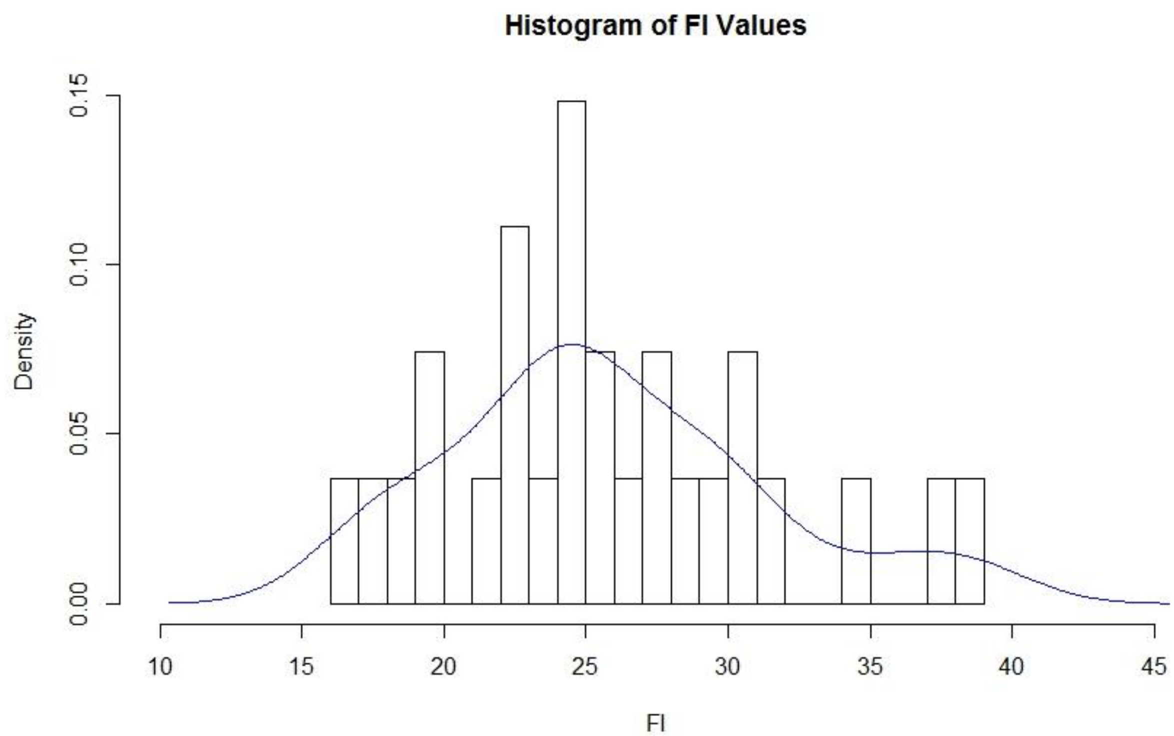
The fracture energy represents the total energy input required to expand a macro crack. The post-peak slope calculated for the Flexibility Index relates to the crack propagation speed. Thus, the mixtures need high fracture energy to prevent the initiation of cracking while keeping a low post-peak slope after the crack is formed. Slow propagation is favored for improving fatigue life of the pavement and preventing quick damage to the entire pavement structure. In other words, high FI is desirable and is achieved when the numerator (G_f) is high while the denominator (slope) is low. Two different materials with a similar amount of fracture energy would be discriminated by the slope of the post-peak curve. Technically the ability of the material to maintain strength after crack initiation differentiates between a flexible behavior and a brittle one. Lab A performed all three sets of tests with a loading rate of 50 mm/min to evaluate the repeatability of the FI values in one Lab. In order to evaluate the effect of loading rate on the post-peak slope, labs B and C also performed one set of tests at a loading rate of 5 mm/min. Finally, to understand the variability of FI, response parameters of fracture energy (FE), Flexibility Index (FI), peak load (PL) and post-peak slope were also investigated.

5.3 Validation of Normality of SBC Sample

Asphalt concrete can be described as a composite viscoelastic material having a spatial disorder with no microstructural periodicity [25]. Therefore, results from the SCB tests between samples will have some inherent variability that needs to be taken into consideration when reporting the results. This work is based on the hypothesis that the FI values measured from the SCB tests are normally distributed. However, normality has never been evaluated, thus the first part of this study evaluated such hypothesis.

To evaluate the normal distribution of FI values, twenty-eight samples as described in Section 5.2 were tested at a loading rate of 50 mm/min and used for the analysis. Given the consistency in sample preparation and the randomization between labs, the 28 samples can be considered to come from the same population and the sample size is large enough to consider the statistical analysis reliable. The normal distribution of FI values was evaluated using a normalized histogram and the normal Q-Q plots (Figure 5-2). The Y-axis for the histogram shows normalized density of the data (percent of occurrence) with a general bell-shaped curve. From the data, normal Q-Q plots were developed by calculating the values of a standard normal quantile (z-scores) of each data and plotting to scatter graphics of FI values data versus z-scores. If scatter points appear to roughly describe a line, then it is reasonable to think the data are normally distributed.

Some minor skewness is expected from the sample distribution. The value of index for skewness (Pearson's coefficient of skewness) is 0.49 which lies down between the accepted values of -0.50 and +0.50 [28]; this indicates that the distribution shown in the histogram has a generally accepted level of skewness to be considered approximately symmetric [27]. The excess kurtosis of eruption duration is -0.25 which indicates that eruption duration distribution is relatively platykurtic (having negative kurtosis and very thin tails compared to the normal distribution). This is consistent with the fact that the histogram is not perfectly bell-shaped. The peak is a bit shallower than the peak of a normal distribution [28]. The R-squared value of the Q-Q plot is high and most of the points fall approximately along the regression line. Thus, it is reasonable to assume that the FI values have a normal distribution and the central limit theorem can be used stating that the distribution of the curve can be considered approximately normal.



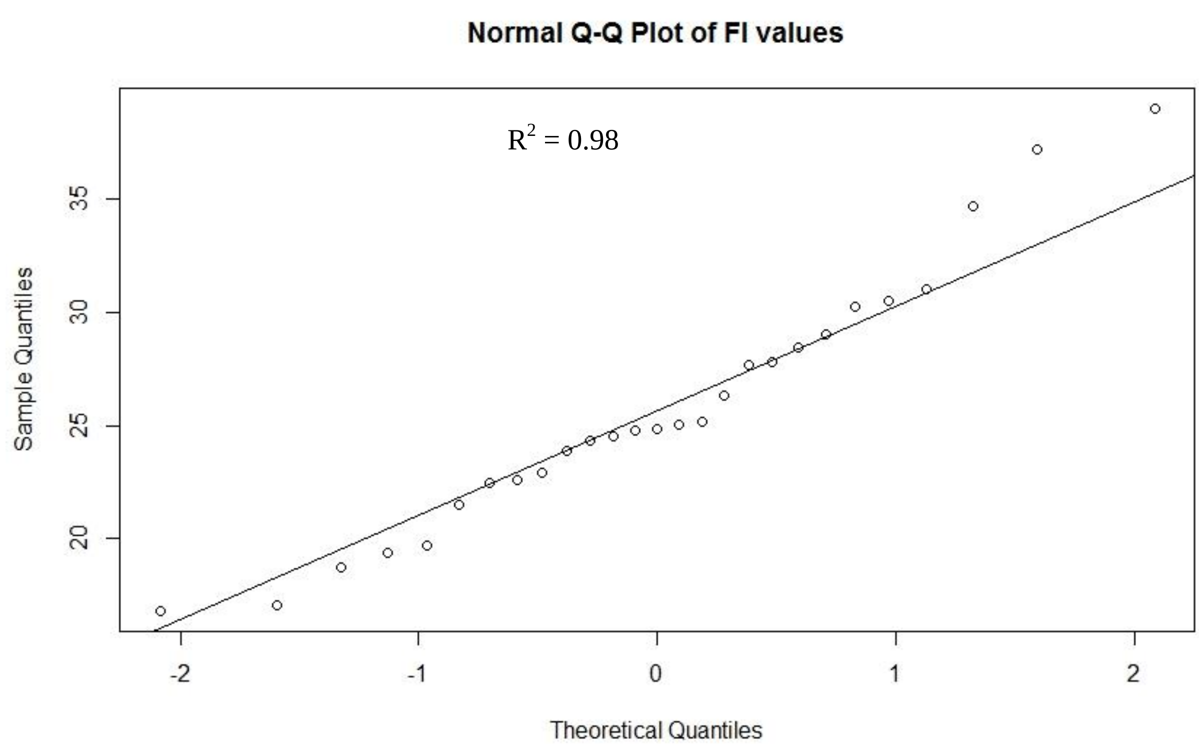


Figure 5-2 Normal probability plots and histograms of FI values

5.4 Analysis of Laboratory Results

Once normality of the data has been addressed, calculating an average of the FI from one gyratory cylinder was evaluated. AASHTO TP124-16 does not describe how to average the results of SCB samples cut from one gyratory cylinder. The authors suggest three possible methods.

Method I: The simplest way is to get an average of four FI values representing four SCB samples obtained from a gyratory puck. If the data is taken from large samples, outliers should not have any considerable leverage or impact. But, since this is a small sample, there is a chance of having values that are outliers. It is noted that outlying values are not necessarily "wrong" but can exist due to the inherent material variability. Outliers should be rare, and if not, the results are not considered reliable.

Method II: The classical approach to screen outlying observations is to use the standard deviation (SD) method. This is defined as 2-SD Method ($x \pm 2$ SD), where the mean is the sample mean and SD is the sample standard deviation. The observations outside these intervals may be considered as outliers and excluded from the analysis [29]. Obviously for this to be reliable, a good estimate of the SD is needed.

Method III: In fracture testing, smaller samples will break at higher stresses than larger samples. An alternative averaging approach would then be to discard the highest value among the four SCB samples taken from one puck and average the remaining three [29, 30].

Table 5-1 presents the daily average from all labs based on the three described methods. For the most part, the standard deviation decreases when methods II or III are used and the calculated mean decreases by 3% when using Method II and by 8% when using Method III. Such decrease in the standard deviation should be viewed with some caution as Methods II and III only use 3 samples resulting in a larger confidence interval (i.e., greater uncertainty). It is noted that some studies have recommended a minimum of six SCB samples to get a reliable FI value [31]. Given the limited data, the authors suggest that, unless a specific reason is found (e.g., known testing errors), the average from all samples is used. Obviously, the more samples available, the more reliable the results will be; yet this must be balanced against the added material and work required to prepare more samples.

Table 5-1 Averaged Data for all Three Laboratories

Lab	Day	Loading Rate (mm/min)	FI mean and Std. Deviation					
			I[1]	S.D.	II[1]	S.D.	III[1]	S.D.
A	1	50	27.5	4.2	29.5*	1.5	26.5*	4.5
	2	50	23.6	3.8	23.6	3.8	21.9*	2.3
	3	50	28.8	9.2	24.2*	1.5	24.2*	1.5
B	1	50	23.3	5.0	25.4*	3.2	21.3*	4.0
	2	50	23.2	4.2	23.2	4.2	22.1*	4.4
	3	5	28.4	7.5	24.7*	1.6	24.7*	1.6
C	1	50	30.9	9.0	30.9	9.0	28.2*	8.8
	2	5	25.0	6.1	25.0	6.1	25.0	6.1
	3	50	27.3	7.0	27.3	7.0	24.8*	6.2

1 Averaging method and standard deviation   * Based on three data points

Table 5-2 shows the analysis of variance (ANOVA) on FI values at different days for Lab A. The results show that there is no significant difference in results between different days. This gives confidence that the SCB test is repeatable. The same results were obtained for both Lab B and C, but omitted for space.

Table 5-2 ANOVA of FI values from Laboratory A

ANOVA: Single Factor

SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Day 1	4	110.09	27.52	17.63		
Day 2	4	94.3	23.58	14.22		
Day 3	4	115.14	28.78	45.42		
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	59.09	2	29.54	0.75	0.49	4.25
Within Groups	351.82	9	39.09			
Total	410.92	11				

Based on the results from Table 5-2, the data from each lab at each loading rate regardless of the day it was tested is analyzed and shown in Table 5-3. One data point from Lab A was eliminated as it was considered an outlying observation (FI value twice as high as the others). It is observed that peak load and fracture energy are noticeably lower at a rate of 5 mm/min as compared to the values at 50 mm/min; however, the same trend is not observed for the FI of this material which leads to the conclusion that the post fracture slope is lower (more ductile failure). It is also observed that the coefficient of variation of the values at the slower rate is higher than the value at the standard rate.

Table 5-3 Combined Test Results

		FI[1]	**FE**[2] **N·m**	**Peak Load kN**	**FE/PL**	**Disp. at Peak mm**
Lab A	n	11				
50 mm/min	**Mean**	**25.2**	**2061**	**1.96**	**1053**	**1.58**
	Std. Dev	3.68	132.09	0.11	70.14	0.14
	Coeff. Var	15%	6%	6%	7%	9%
Lab B	n	8				
50 mm/min	**Mean**	**23.2**	**2257**	**2.21**	**1021**	**1.70**
	Std. Dev	4.28	248.89	0.25	46.63	0.11
	Coeff. Var	18%	11%	11%	5%	6%
Lab B	n	4				
5 mm/min	**Mean**	**28.4**	**1184**	**1.10**	**1077**	**1.68**
	Std. Dev	7.52	179.00	0.14	110.63	0.10
	Coeff. Var	27%	15%	13%	10%	6%
Lab C	n	8				
50 mm/min	**Mean**	**29.1**	**2529**	**2.24**	**1130**	**1.68**
	Std Dev	7.70	315.32	0.16	102.13	0.21
	Coeff. Var	26%	12%	7%	9%	12%
Lab C	n	3				
5 mm/min	**Mean**	**22.5**	**1134**	**1.13**	**999**	**1.67**
	Std Dev	6.22	42.00	0.06	46.64	0.12
	Coeff. Var	28%	4%	5%	5%	7%
All Labs	n	27				
50 mm/min	**Mean**	**25.8**	**2258**	**2.12**	**1067**	**1.64**
	Std Dev	5.63	298.62	0.21	85.13	0.16
	Coeff Var	22%	13%	10%	8%	10%

[1] Flexibility Index [2] Fracture Energy

When combining all the results at 50 mm/min, the mean value of FI is found to be 25.8, which is a relatively high value in comparison to other values reported in the literature [2, 31]. With a standard deviation of 5.63, the coefficient of variation is calculated to be 22%. It is observed in Table 5-3 that the coefficient of variation of the fracture energy and of the peak load at 13% and 10%, respectively, are considerably lower than 22% which means that a significant portion of the variability in the FI comes from the value of |m| slope. It is not known if the reason for the higher variability in the |m| slope comes from the method by which the inflection point is determined or if it represents actual variability in the post cracking behavior of the material. More research is needed to understand this issue.

A t-test on the results shows that the samples tested at laboratory A and laboratory B are gathered from the same source of material but not samples from laboratory C. This means that the SCB test is reproducible between two labs. A review of laboratory C's procedure was done and the systematic error was addressed and corrected for the field phase of the study.

5.4.1 Population Variability

Given that all samples were fabricated in the same manner, it can be argued that the standard deviation, and thus the coefficient of variation of the population, is close to the value obtained for all labs at 50 mm/min (n=27), a value of 22% (5.63/25.8). Using this value and Equation 5, a confidence interval can be constructed to determine the range of values that can be expected from testing different number of samples.

$$\mu = \bar{x} \pm t_{\alpha/2} \frac{\sigma}{\sqrt{n-1}} = \bar{x} \pm t_{\alpha/2} \frac{0.22\bar{x}}{\sqrt{n-1}} \qquad \text{Equation 5}$$

In this equation μ is the mean of the population, $\bar{x}$ is the average of the test, $t_{\alpha/2}$ is the t-value coefficient based on probability of 1-α for a two-tailed distribution, σ is the standard deviation of the population (22% of the mean as previously discussed), and n is the number of the samples tested. As the equation shows, the range of expected values decreases as the number of samples increase; in other words, the uncertainty in the results decreases with more tests. Using Equation 5, the values shown in Table 5-4 are determined.

Table 5-4 Range of Values for Different Number of Tests Based on Standard Deviation

Number of Tests	*Degrees of freedom*	*2-tail t-value $\alpha = 0.05$*	*Range of values $\bar{x} \pm$*
3	2	4.303	66.9%
4	3	3.182	40.4%
6	5	2.571	25.3%
8	7	2.365	19.7%
15	14	2.145	12.6%
20	19	2.093	10.6%

The results presented in Table 5-4 indicate that at least eight samples need to be tested to decrease the range of expected values below 20%. This is a higher number than the six samples that some studies have recommended; however, from a practical standpoint, both require compaction of two gyratory cylinders so testing eight samples adds no significant extra work.

5.5 Effect of Loading Rate

The results from Table 5-3 show that there is little advantage in reducing the loading rate to 5 mm/min. The lower loading rate results in higher FI in Lab B but not in Lab C. The peak load decreases but the fracture energy increases as the rate decreases. The standard deviation increased at the lower loading rate. However, during testing, it was observed that, at a loading rate of 50 mm/min, some aggregates were fractured, a behavior that is not supported by field observations and giving a reason to evaluate slower loading rates, even though a lower rate might result in a more dominant plastic zone. Given the lack of clear trends and the small sample size, more testing at different rates is recommended before any recommendations can be made. This topic will be investigated in more detail in Section 6.4.

5.6 Summary

In this study, the Flexibility Index and fracture properties of asphalt mixtures produced in the lab were analyzed. Based on statistical evaluations, the following observations were made:

- The study showed that even though there is a slight skewness towards the high end, the results from the test can be considered normally distributed; thus descriptive statistics can be used.
- The comparison of results between 3 labs indicated that, while the results are reproducible within each lab on repeated days, it is possible that a bias is introduced by a lab. Thus, it is important to verify on a regular basis that all labs are getting statistically similar results.
- The study revealed that at least 8 samples should be tested to obtain an average that represents the actual value within 20%. This requires compaction of 2 gyratory pucks.
- A coefficient of variation (CV) between 20% and 30% was observed for samples cut from one puck. When comparing the average of four samples cut from one puck to another similar puck, the difference in results was around 11%.
- There was little advantage found in performing the tests at a lower loading rate; however, more testing is recommended.

6.0 EVALUATION OF FLEXIBILITY INDEX USING FIELD MATERIAL

6.1 Description

As was described in Section 3, representative materials from across the state of Utah, based on both UDOT and non-UDOT projects, were collected from seven different projects. The material was distributed to three different labs where each of the labs compacted, cut, and tested the samples based on the procedures described in AASHTO TP124-16. Unfortunately, due to variations in field collected materials, not all samples obtained were within the limits, in terms of air voids or cutting geometry, specified in the AASHTO specifications. Given the limited availability of materials, all samples were tested and the all the data is reported.

6.2 Results

Testing of the material was done following the procedures described in AASHTO TP124-16 with minor exceptions discussed in this section. Each of the three participant labs prepared two gyratory compacted cylinders for each mixture at each condition. As was discussed in Section 5, four semi-circular samples can be obtained from one gyratory-produced cylinder, thus the two compacted cylinders resulted in 8 samples for each mixture at the two conditions (plant and field laydown). As was recommended in Section 5.4, at least eight samples should be tested for a reliable mean; however, due to limitations in materials, this was not always possible. The overall methodology for the test was described in Section 5.2.

All of the labs tested the samples at 50 mm/min; and, to evaluate the effect of loading rate as a follow up of the discussion in Section 5.5, a partial factorial experiment was conducted and samples were tested at 15 mm/min, 5 mm/min, and 0.5 mm/min.

Table 6-1 shows the average FI of each gyratory sample (i.e., puck) tested (average of 4 tests), the coefficient of variation (standard deviation divided by the mean) of all the samples tested, and the mean of the results for each mixture at each of the testing labs.

Table 6-1 Results of FI at 50 mm/min

		Lab A		Lab B		Lab C	
		Plant	**Field**	**Plant**	**Field**	**Plant**	**Field**
UT-01	Puck 1	6.7	4.6	9.6	7.4	5.8	6.9
	Puck 2	5.1	7.2	4.6	8.2	-	-
	Coeff Var[1]	**31%**	**39%**	**38%**	**17%**	**14%**	**26%**
	Average	**5.9**	**5.9**	**7.1**	**7.8**	**5.8**	**6.9**
UT-02	Puck 1	5.5	3.7	3.9	3.6	3.1	2.1
	Puck 2	4.3	3.0	4.0	3.4	3.4	-
	Coeff Var	**29%**	**24%**	**24%**	**16%**	**25%**	**38%**
	Average	**4.9**	**3.4**	**4.0**	**3.5**	**3.3**	**2.1**
UT-03	Puck 1	12.0	8.7	11.7	8.9	7.0	9.3
	Puck 2	4.6	-	7.7	12.8	-	-
	Puck 3	-	-	-	13.5	-	-
	Coeff Var	**20%**	**27%**	**32%**	**30%**	**24%**	**29%**
	Average	**8.3**	**8.7**	**9.7**	**11.7**	**7.0**	**9.3**
UT-04	Puck 1	15.3	10.0	8.7	10.6	13.4	9.5
	Puck 2	8.4	7.4	9.5	9.5	-	-
	Coeff Var	**38%**	**27%**	**20%**	**27%**	**32%**	**40%**
	Average	**11.8**	**8.7**	**9.1**	**10.1**	**13.4**	**9.5**
UT-05	Puck 1	3.8	4.5	11.3	5.2	6.8	4.8
	Puck 2	7.8	9.4	11.8	4.8	8.8	-
	Coeff Var	**39%**	**40%**	**19%**	**19%**	**13%**	**24%**
	Average	**5.8**	**7.0**	**11.6**	**5.0**	**7.8**	**4.8**
UT-06	Puck 1	2.9	3.3	7.8	6.0	2.9	2.1
	Puck 2	3.2	4.1	3.4	6.5	3.2	2.4
	Coeff Var	**23%**	**18%**	**47%**	**30%**	**20%**	**20%**
	Average	**3.0**	**3.7**	**5.6**	**6.2**	**3.0**	**2.3**
UT-07	Puck 1	14.3	10.1	18.8	18.8	8.4	8.4
	Puck 2	9.0	15.8	22.5	17.8	-	-
	Coeff Var	**28%**	**29%**	**22%**	**24%**	**29%**	**28%**
	Average	**11.6**	**12.9**	**20.6**	**18.3**	**8.4**	**8.4**

1 Coefficient of Variation (standard deviation divided by the mean) for all the samples tested. Four tests are obtained for each puck resulting in 8 samples.

As can be seen in Table 6-1, it is not unusual to have a significant difference in FI between the first gyratory puck and the second gyratory puck. Most previous work, including that described in Section 5, indicate that more than one gyratory puck should be compacted and testing should be done on as many individual SCB samples as possible to obtain a reliable mean.

These data supports such statement. The data also shows that the coefficient of variation of all results is often greater than 20% and that there are differences in the results obtained between all three labs that are greater than the expected variability. Mixtures from UT-06 and UT-07 show particularly large differences between labs. Also, for the plant mixes, with the exception of UT-02, the results from Lab A and Lab C are closer to each other than to Lab B. In 5 out of the 7 mixtures obtained at the plant and in 6 out of the 7 mixtures obtained at laydown, Lab B had higher FI values than the other labs. As was mentioned in Section 3.5, Lab B had more material and was able to compact more samples that were within the target air voids and dimensions. It is possible that the bias seen on Lab B values is actually the result of more control of the samples tested. This will be discussed in Section 6.5.

In an effort to improve the variability of the results, Method III of averaging, described in Section 5.4, was used. In this method, the highest value from each gyratory sample (puck) is eliminated and the resultant 6 samples are averaged for Labs A and B. This was not done for Lab C to ensure that at least 6 data points are used to obtain a reliable average. The results are shown on Table 6-2.

Table 6-2 FI Results without the Highest Value

		Lab A		Lab B	
		Plant	**Field**	**Plant**	**Field**
UT-01	Average	5.3	5.1	7.0	7.5
	Coeff Var	31%	40%	39%	19%
UT-02	Average	4.4	3.1	3.6	3.2
	Coeff Var	26%	23%	14%	11%
UT-03	Average	11	-*	9	10
	Coeff Var	20%	-*	33%	26%
UT-04	Average	10.8	7.9	8.4	9.1
	Coeff Var	33%	22%	17%	25%
UT-05	Average	5.5	5.7	10.5	6.5
	Coeff Var	38%	39%	14%	17%
UT-06	Average	4.1	3.6	5.1	5.3
	Coeff Var	17%	18%	49%	13%
UT-07	Average	11.0	12.2	18.7	16.3
	Coeff Var	29%	33%	6%	14%

*Less than 6 samples available

The data in Table 6-2 shows that, even after using Method III, there is significant variability in the results. As was discussed in Section 5.4, it is hypothesized that most of the variability comes from the calculation of the slope. To verify such claim, the variability of the individual parameters that are used to calculate the Flexibility Index (FI), namely Fracture Energy and slope (recall Equation 4) were determined for the data in Lab A for the 50 mm/min data. These results are summarized in Table 6-3

Table 6-3 Coefficient of Variation for Different Parameters in Lab A

		Coefficient of Variation[1], %		
		Fracture Energy	Slope	Flexibility Index
UT-01	**Plant**	8.5	28.0	31.1
	Field	19.9	22.4	38.6
UT-02	**Plant**	13.1	21.7	29.0
	Field	15.0	15.3	23.8
UT-03	**Plant**	11.1	19.0	19.9
	Field	16.6	13.8	27.0
UT-04	**Plant**	9.0	41.5	37.7
	Field	7.2	27.1	26.8
UT-05	**Plant**	6.0	42.2	38.7
	Field	9.1	43.8	39.7
UT-06	**Plant**	8.4	18.8	22.6
	Field	12.0	11.5	17.5
UT-07	**Plant**	13.3	19.3	27.8
	Field	9.7	34.2	28.7

1. the coefficient of variation is the standard deviation divided by the mean based on 8 samples tested

As can be seen in Table 6-3, the coefficient of variation of the fracture energy is, on average, less than 12% and, in 8 out of 14 sections, less than 10%. In contrasts, the coefficient of variation of the slope is, on average, 26% and in only 2 out of 14 sections the value is less than 15%. The Flexibility Index, on average, has coefficient of variation of 30% which is higher than expected for a within-lab specification test. More research is needed to determine if an alternative parameter can result in lower test variability while still capturing the desired flexibility of the material.

Even though the data shows that there are variability issues and differences between testing labs that need to be resolved, there is consistency when it comes to identifying mixtures that might result in poor performance. The results from all labs show that plant mixtures from projects UT-02 and UT-06 are likely to have poor intermediate-temperature performance since both have low FI values (below 6). In the same manner, all labs show that plant mixtures from projects UT-03, UT-04, and UT-07 have FI values above 7 and, thus, better intermediate-temperature performance would be expected. For other the two sections (UT-01, and UT-05), the results are inconclusive since some labs got FI values below 6 and some have FI values above 6. These results are shown graphically in Figure 6-1.

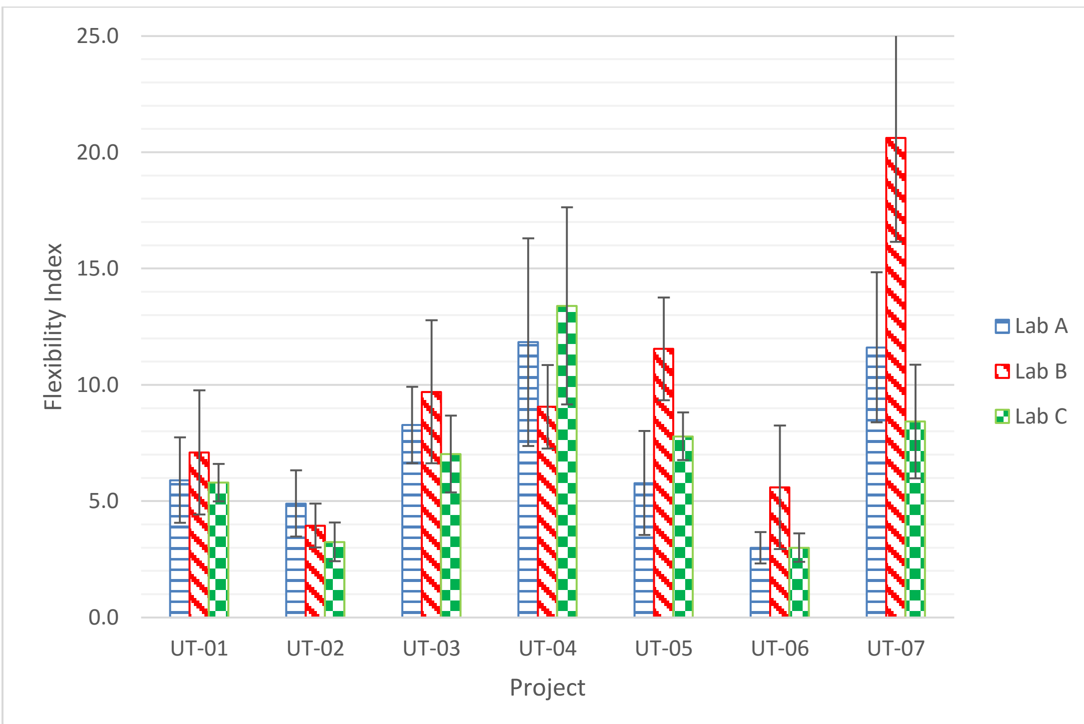


Figure 6-1 Average Flexibility Index of Plant Mixtures at all Three Labs.
Error bars represent ± one standard deviation

6.2.1 Comparison to Low-Temperature Results

A comparison between the performance predictions at low temperature (Table 4-6 for PG XX-34) and the predictions at intermediate temperature (FI <6) result in some commonalities. Both tests are in agreement that material from sections UT-03, UT-04, and UT-07 would not be expected to crack while materials from sections UT-01, UT-02, and UT-06 might not have good cracking performance. There is no agreement with material from UT-05 given the difference in FI obtained between labs.

6.3 Effect of Short-Term Aging

As was explained in Section 3.3, material was collected at two locations, at the plant and at laydown in the field. Due to storage and transportation to the site, the material collected at laydown (i.e., field material) is the same material collected at the plant but in a short-term aged condition. The significance of change in FI from short-term aging can be evaluated by comparing the test results.

Before analysis of the short-term aging of mixtures is discussed, given the variability of the test results, it is important to understand what values can be considered a significant difference. Assuming an FI threshold value of 6 and a variability of 20%, a change in FI of 1.2 can be considered to be not significant. Using this as reference, the data indicates that in mixtures from sections UT-01, UT-02, and UT-06 there is no significant aging (i.e., the difference in FI between the plant and laydown is less than 1.2). As a reference, mixtures from sections UT-01 and UT-02 also showed no significant short-term aging in the low-temperature results (see Section 4.4.1); however, the mixture from section UT-06 showed significant aging at low temperatures but no change in FI.

For mixtures from sections UT-04 and UT-05 there is a decrease in FI in two out of the three labs. It should be recalled that the effect of aging on mixtures from these two sections was not detected in the BBR tests; in fact, the aging results from sections UT-04 and UT-05 were an anomaly in regards to aging. The mixture from UT-05 was stored in a silo for 17 hours so some aging is expected but only detected in the FI value.

The FI from mixtures obtained in projects UT-03 and UT-07 show significant changes but contradictory FI results from different labs regarding aging. At low temperatures, mixtures from UT-03 show no significant aging while mixtures from UT-07 show significant aging. Based on mixed results, no conclusions are reached for these two sections. These results are shown graphically in Figure 6-2.

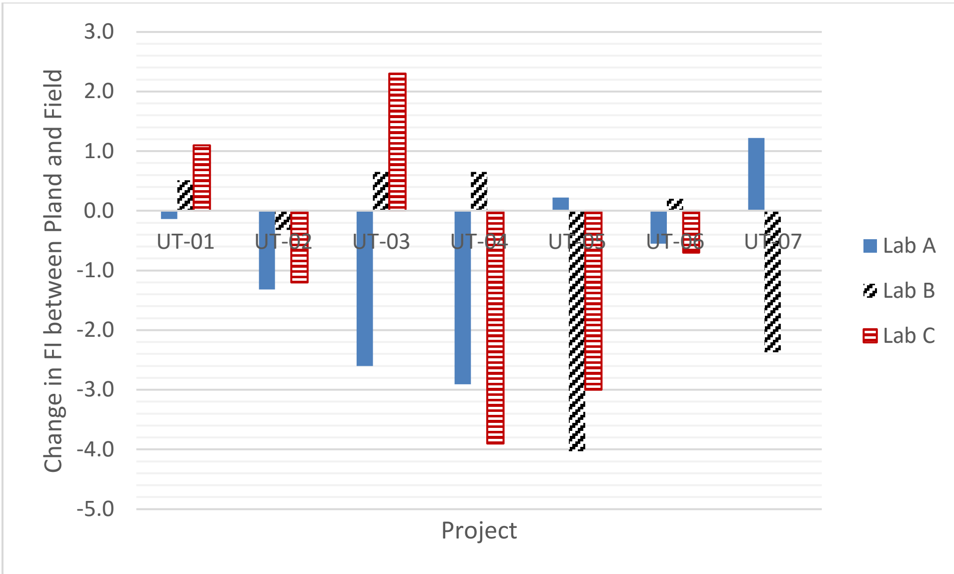


Figure 6-2 Change in FI between Plant and Laydown

6.4 Effect of Loading Rate

As was discussed in Section 5.5, one of the questions regarding the determination of the FI is whether the rate of loading is adequate. Anecdotal evidence suggests that at 50 mm/min the crack propagation might be affected by broken aggregates. Figure 6-3 shows a picture illustrating such cases for a particular mixture tested at 50 mm/min and 5 mm/min loading rates. The picture from the test at the higher loading rate shows many broken aggregates when compared to the picture from the test at the lower loading rate. The energy required to break aggregates might result in added variability.

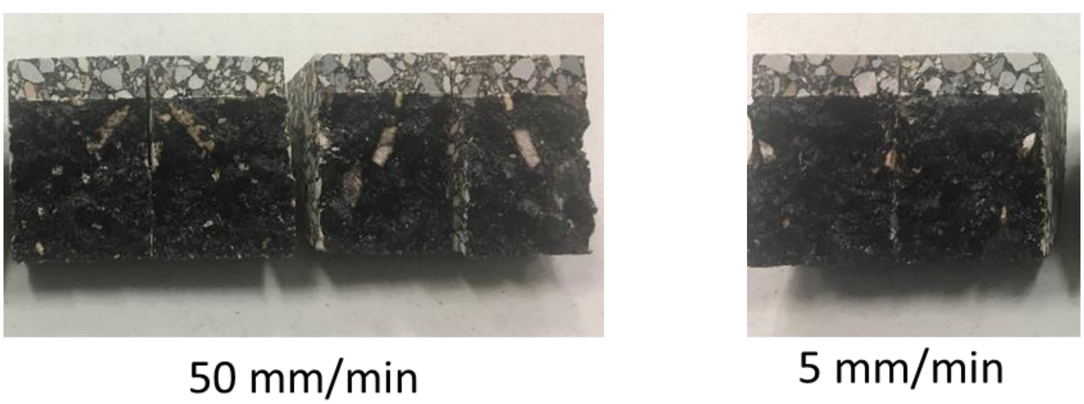


Figure 6-3 Picture showing more broken aggregates at a faster loading rate

To investigate if there is any validity to the effects of different loading rates, a partial factorial experiment was devised in which Lab A tested at 5 mm/min, Lab B tested at 15 mm/min and 0.5 mm/min, and Lab C tested at 15 mm/min and 5 mm/min. This ensured that the results at 15 mm/min and 5 mm/min are duplicated in at least 2 labs.

The results for different loading rates are shown in Table 6-4. These values can be compared to the standard loading rate of 50 mm/min shown in Table 6-1.

Table 6-4 Average FI Results at Various Speeds

		Lab A		**Lab B**		**Lab C**		**Difference in Results[1]**	
	Speed	Plant	Field	Plant	Field	Plant	Field	Plant	Field
UT-01	15	-	-	15.3	10.7	12.5	11.6	18%	-8%
	5	14.0	9.5	-	-	17.0	11.3	-21%	-20%
	0.5	-	-	13.3	20.8	-	-	-	-
UT-02	15	-	-	3.6	2.5	3.4	3.2	6%	-25%
	5	-	4.3	-	-	5.4	3.6	-	17%
	0.5	-	-	10.2	4.5	-	-	-	-
UT-03	15	-	-	10.1	13.5	10.2	9.2	-1%	32%
	5	-	-	-	-	12.3	10.4	-	-
	0.5	-	-	-	12.1	-	-	-	-
UT-04	15	-	-	9.0	11.0	10.0	9.0	-11%	18%
	5	20.5	10.1	-	-	8.9	8.7	57%	13%
	0.5	-	-	11.7	8.9	-	-	-	-
UT-05	15	-	-	11.8	4.6	11.1	6.4	6%	-41%
	5	5.4	9.5	-	-	-	5.7	-	40%
	0.5	-	-	18.5	8.8	-	-	-	-
UT-06	15	-	3.9	7.5	5.6	-	-	-	-
	5	7.3	6.0	-	-	5.5	6.4	25%	-6%
	0.5	-	-	8.5	9.4	-	-	-	-
UT-07	15	-	-	24.5	14.8	11.6	11.3	53%	24%
	5	21.5	26.0	-	-	11.7	10.4	45%	60%
	0.5	-	-	27.3	21.3	-	-	-	-

[1] First minus second lab divided by first lab

As seen in Table 6-4, the difference in results from any two labs, a measure of the between-lab variability, was still high for different loading rates. For the 15 mm/min loading rate, 5 out of 12 results show a difference between labs greater than 20% with of them 2 being greater than 40%. For the 5 mm/min the difference was worse with 7 out of 10 results showing a difference between two labs greater than 20% with 4 out of them being greater than 40%. As a reference, in Table 6-1, the difference between Labs A and C, the two closest to each other, was greater than 20% in 7 out of 14 results. In terms of the coefficient of variation when the tests from all labs are averaged, there was no significant improvement when compared to the standard loading rate of 50 mm/min.

The data from all labs was averaged, and from this value the average FI obtained at 50 mm/min was subtracted. This approach gives a general trend while considering all of the data; however, it does not account for some of the variability in the results or the fact that some values are based on more samples than others (see discussion in Section 5.4.1). Given those limitations, a graphical representation of the results is shown in Figure 6-4 for the material collected at the plant.

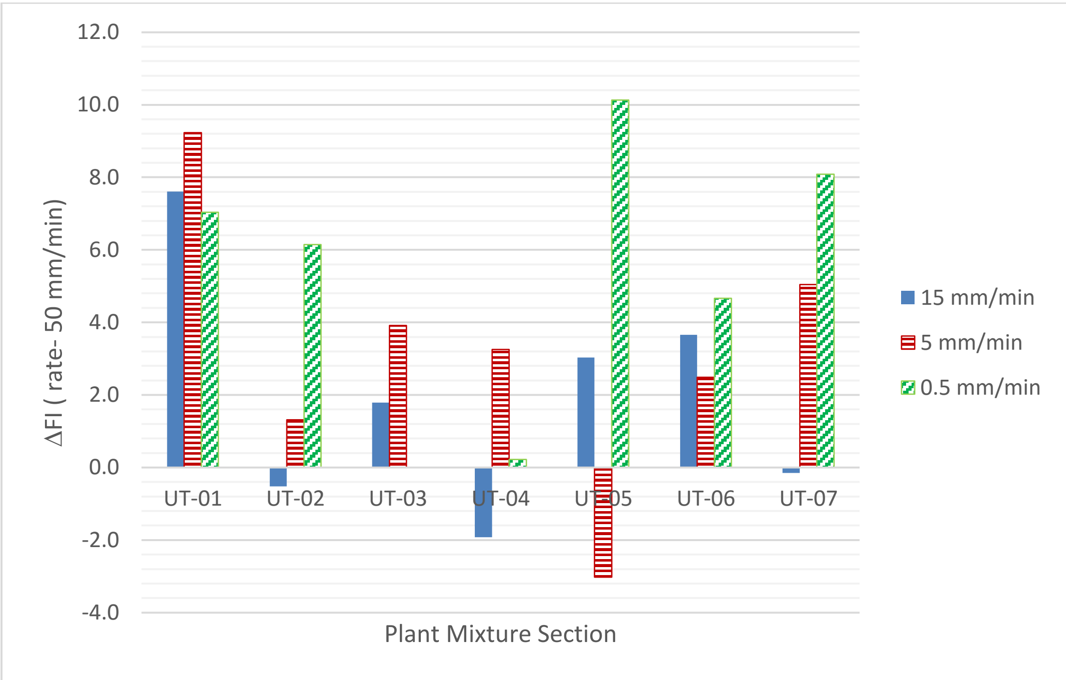


Figure 6-4 Effect of Loading Rate for Mixtures Collected at the Plant

The results in Figure 6-4 show that a higher Flexibility Index is obtained at slower loading rates. In four out of the seven sections tested, the highest FI is obtained at the lowest loading rate (0.5 mm/min) while in three of them the highest value is obtained at 5 mm/min.

Comparing the results shown in Table 6-4 to those shown in Table 6-1 would lead to the same conclusion: mixtures from section UT-07 should have the best performance and mixtures from section UT-02 and UT-06 would likely have the worst performance. The results for the other sections are not as clear since different rates result in different relative values. To illustrate these changes, the Flexibility Index based on the average of the values obtained at all labs for plant mixtures is presented in Table 6-5 along with the relative ranking of the mixtures based on the expected performance. Arrows in the table connect each section's results.

Table 6-5 Performance Ranking of Plant Mixtures at Different Speeds

Rank[1]	**Loading Rate**			
	50 mm/min	**15 mm/min**	**5 mm/min**	**0.5 mm/min**
1	UT-07 FI = 13.6	UT-07 FI = 18.1	UT-07 FI = 16.6	UT-07 FI = 27.3
2	UT-04 FI = 11.4	UT-01 FI = 13.9	UT-01 FI = 15.5	UT-05 FI = 18.5
3	UT-05 FI = 8.4	UT-05 FI = 11.4	UT-04 FI = 14.7	UT-01 FI = 13.3
4	UT-03 FI = 8.3	UT-03 FI = 10.1	UT-03 FI = 12.3	
5	UT-01 FI = 6.3	UT-04 FI = 9.5	UT-06 FI = 6.4	UT-04 FI = 11.7
6	UT-02 FI = 4.0	UT-06 FI = 7.5	UT-05 FI = 5.4	UT-02 FI = 10.2
7	UT-06 FI = 3.9	UT-02 FI = 3.5	UT-02 FI = 5.4	UT-06 FI = 8.5

[1] Based on expected performance as determined by the FI

Given that there is no field performance data that has been measured for these seven sections, it is not known if any of the rates actually result in performance predictions that better match the field.

6.5 Effect of Sample Preparation

As was mentioned in Section 3.5, all three labs found that producing samples with consistent air voids and dimensional cuts was difficult, and many of the samples did not meet the production standards defined in the AASHTO TP-124 sample preparation procedure. To evaluate if sample preparation had an effect on the results, a comparison was made using the data from the field (laydown) samples tested at 50 mm/min at all three labs. In one case, the data from all different labs was averaged and the coefficient of variation determined for each section. In the other case, those samples that had air voids outside the limits specified, 7 ± 0.5 %, and those samples that had a notch depth outside the 15 ± 0.5 mm limits were eliminated from the analysis. While this reduced the number of samples in one data set, it allows for some relative comparisons of the effects of sample preparation. These results are presented in Table 6-6.

Table 6-6 Comparison of the Effect of Sample Preparation on FI

	Field Laydown Section						
	UT-01	**UT-02**	**UT-03**	**UT-04**	**UT-05**	**UT-06**	**UT-07**
	All data						
n	20	20	20	20	19	24	20
Average FI	6.9	3.1	10.6	9.4	5.7	5.0	14.1
C of V	28.5%	26.1%	28.2%	28.8%	37.7%	37.8%	37.3%
	Only data from samples within specifications						
n	8	15	17	8	3	11	9
Average FI	6.5	3.3	10.2	9.7	4.3	3.3	17.4
C of V	24.5%	22.9%	28.7%	34.4%	14.8%	36.7%	23.1%
	Observed Change						
C of V	-4%	-3%	+1%	+6%	-23%	-1%	-14%
Average	-0.4	+0.2	-0.4	+0.3	-1.4	-1.7	+3.3

The results, shown in Table 6-6, indicate that by eliminating the data from those samples that do not meet specification, in terms of air voids and notch length, the average FI remains essentially the same (< 1.5 change in 5 out of the 7 sections). Also, assuming a limit of 6, no changes in predicted performance is seen. With respect to the variability, as determined by the coefficient of variation, less than a 4% change in the coefficient of variation for 4 out of the 7 sections is seen. In one of the sections (UT-07) there is a 14% decrease in the coefficient of variation when the data from out-of-spec samples is eliminated. As was shown in Figure 6-1, this section has the highest FI of the set but also showed the highest variability. Section UT-05 shows a 23% decrease in the coefficient of variation; however, results are based on only 3 samples. As was discussed in Section 5.4, caution should be used when making observation with less than 8 samples. Data from sections UT-03 and UT-04 show an increase in the coefficient of variation when out-of-spec sample tests are eliminated.

6.6 Summary

The results presented in this section indicate that more work needs to be done to control the variability of the Flexibility Index, both in terms of within-lab variability as well as differences between labs. Even after a trimmed mean approach was used, where the highest value was eliminated, the coefficient of variation was larger than what is desired in a specification test. While it is believed that issues such as control of the air voids after cutting and tolerances for sample preparation (e.g., notch depth) can play a role in the variability of the results, this was not shown to be the case for one lab.

Notwithstanding the variability observed in the Flexibility Index, comparisons of the results obtained at the different labs indicate that the test can consistently predict the extreme expected performers from the different mixtures collected. Based on the literature, an FI limit between 6 and 10 would separate mixtures based on their expected performance. Application of this limit would result in 3 mixtures being eliminated. Furthermore, the predictions are consistent with the results obtained using the BBR.

Regarding the aging that occurred between the plant and laydown, the results indicate that the effects were mixture-specific and not always consistent with the results at low temperatures. For example, mixtures from sections UT-01 and UT-02 showed no aging in both tests while section UT-06 showed aging in the BBR results but no change in FI. Mixtures from sections UT-04 and UT-05 showed clear indications of aging (i.e., a decrease in FI) but were considered an anomaly in the BBR results.

Finally, testing was done at different loading rates of 50 mm/min, 15 mm/min, 5 mm/min and 0.5 mm/min. A small improvement in the differences between labs was seen at 15 mm/min, but there was no clear benefit regarding the coefficient of variation. Furthermore, each loading rate resulted in different performance ranking of the mixtures. Unfortunately, without any performance information available from the field, it is not known if one loading rate is preferred over the other. It is noted, however, that a rate of 0.5 mm/min generates large amounts of data, making the analysis significantly more cumbersome.

7.0 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

7.1 Summary

Seven asphalt mixtures were collected at different plants from the slats and at laydown from the windrow. The mixtures were considered representative of the material produced in the state of Utah. The mixtures were compacted in three different labs and tested at low and intermediate temperatures using the bending beam rheometer (AASHTO TP125-16) and the semi-circular bend test configuration (AASHTO TP124-16). The creep modulus and relaxation capacity were determined at low temperatures, and the Flexibility Index was measured at intermediate temperatures. Both tests provided insight as to the potential performance of the mixtures.

7.2 Low-Temperature Cracking

The study of field-produced mixtures in the state of Utah indicates that three out of the seven mixtures tested (UT-03, UT-04, and UT-07) are expected to have good performance even at the lowest temperature environment of PG XX-34. While all of these mixtures had a creep modulus above 12,000 MPa at the test temperature of -24 ºC, their m-value was above 0.12 indicating good relaxation capacity. The range in the RAP content of these mixtures varied from a low of 10% (UT-07) to as much as 25% (UT-03) indicating that RAP content alone is not a good indicator of expected performance. This supports the notion that the low-temperature performance of the mixture does not depend on a single design parameter (i.e., RAP content, binder grade, etc.) but rather on how all components of the mix combine into a system.

All seven mixtures collected are expected to have good low-temperature performance at the warmer environment of PG XX-22.

In four out of the seven sections tested (UT-02, UT-03, UT-06, and UT-07), there is an increase in modulus and a decrease in m-value between the material collected at the plant and the material collected in the field indicating that short-term aging occurred. In two of the sections tested (UT-01 and UT-05), there was not a clear indication of aging as the results are within the margin of error. Based on previous work, comparison of the short-term aging in the mixtures studied indicates that 3 hours of loose-mix oven aging at the compaction temperature should replicate, for most cases, what occurs in the field.

The within-lab repeatability of the BBR results was usually below 10%. Results for the between-lab repeatability comparison seemed to show a bias in the modulus measurement for one of the labs. No specific cause for this bias was identified, but further repeated testing confirmed the results from previous studies in which the difference between labs was less than 10%.

7.3 Intermediate-Temperature Cracking

The study of laboratory-produced mixtures under consistent compaction and cutting conditions resulted in a coefficient of variation of 22%. If this variability is applied on a confidence interval, the resulting conclusion is that at least 8 tests are needed to obtain a reliable average in FI. This translates into two gyratory compactor cylinders.

The Flexibility Index of mixtures produced in the state of Utah generally ranged from a low value of 3.0 to a high value of 13.5 for plant-produced, unaged material. Those mixtures with the lowest virgin binder content resulted in the lowest FI. Short-term aging resulted in a relatively small decrease in FI for three of the mixtures tested, while in the remaining four, short-term aging resulted in a decrease in FI of up to four.

A study of different loading rates indicated that slower rates will result in higher FI values. Different rates also resulted in different rankings for the same mixture. However, since no performance data is available and no reference exists regarding the FI value at other rates, more studies need to be conducted before a limit or threshold can be adopted. Therefore, no conclusion can be reached regarding a preferred loading rate at this time.

Given the large variability observed in the results, it is not believed that the Flexibility Index parameter can be used to accurately rank the expected performance of different mixtures. In other words, an asphalt mixture with an FI of 11, such as UT-04, might not necessarily have better performance than another mixture with an FI of 8, such as UT-05. It is possible, however, that once a threshold is established, the test can be used to identify mixtures that are susceptible to early fatigue cracking; a pass-fail type test. For example, if a threshold of 6, as suggested by other states, is found adequate, then UT-02 and UT-06 should not be placed on the road. Actual field performance is needed before a determination can be made.

Finally, it was observed that sample preparation requires significant effort in terms of materials, compaction, and cutting. More research is needed to determine how the sample preparation affected the large variability observed in the test.

7.4 Relation between Performance Tests

Each of the tests evaluated as part of this project represent a potential source of information regarding the expected performance of asphalt mixtures. The selection (or rejection) of a given mixture should not be seen as a single index decision but a continuum that is dependent on the environment. Table 7-1 shows a predicted performance comparison of all seven mixtures tested at various environments. In the table an “X” means poor performance would be expected, and a “thumbs up” means good performance would be expected.

Table 7-1 Performance Comparison of Different Mixtures

	Lowest Temp[1] PGXX-34	**Low Temp[1] PG XX-28**	**Intermediate[2] Temperature**
UT-01	❌	❌	❌
UT-02	❌	❌	❌
UT-03	👍	👍	👍
UT-04	👍	👍	👍
UT-05	❌	❌	👍
UT-06	❌	❌	❌
UT-07	👍	👍	👍

[1] BBR results, [2] SCB FI results

Table 7-1 shows that there is some commonality in the mixtures deemed adequate at different environments using the different tests, even though intermediate-temperature predictions are still evolving. The threshold for the Flexibility Index will continue to evolve until field performance observations become available. Nonetheless, three of the seven mixtures tested are predicted to have adequate performance at both low and intermediate temperatures. It is noted that UT-03, UT-04, and UT-07 all have different mixture design and RAP content ranging from 10% up to 25%, indicating that mix performance is not dependent on a single parameter but rather in the proper formulation of the mixture as a system. It is also noted that those three mixtures come from 3 different UDOT administrative regions, indicating that good performing mixtures can be designed across the state.

7.5 Conclusions

After extensive testing of asphalt mixtures collected from seven different pavement sites across the state of Utah, the following conclusions are reached.

1. The low-temperature limits proposed as part of a previous study will allow evaluation of the expected performance of asphalt mixtures at specific low-temperature environments. While most of the mixtures produced have a relatively high creep modulus at the intended environment (creep modulus >12,000 MPa), their relatively high relaxation capacity (m-value >0.12) should result in good performance. These predictions are based on the mixture as a system and are not based on individual parameters such as neat asphalt binder grade or RAP content.
2. Variability in the within-lab and between-lab results at intermediate temperature (FI) continues to be a problem. While sample preparation was a challenge and might have contributed to some of the observed variability, the actual source of the high variability remains unknown. Based on a controlled study, it was found that least 8 samples should be tested to obtain an average that represents the actual value within 20%. This requires compaction of 2 gyratory pucks.
3. The aging that occurs between the plant and laydown is mixture-specific and is not always consistent with the results at different temperatures. Based on low-temperature testing, the current practice of loose-mix aging for 2 to 4 hours is adequate to simulate the changes that are observed in the field.
4. A small improvement in the differences between labs was seen when FI testing was done at 15 mm/min, but there was no clear benefit regarding the coefficient of variation. Furthermore, each loading rate resulted in different performance ranking of the mixtures.

Unfortunately, without any performance information available from the field, it is not known if one loading rate is preferred over the other

5. Notwithstanding the large coefficient of variation in the data, the Flexibility Index test can predict the extreme expected performers out of the different mixtures collected. Asphalt mixtures sampled at the plant can be expected to have an FI generally between 3.0 and 20.0. The material sampled at laydown can be expected to have an FI somewhere between 2.1 and 18.5. Based on the literature, an FI limit between 6 and 10 would separate mixtures based on their expected performance. Application of this limit would result in 3 out of the 7 mixtures being eliminated. These predictions are consistent with the results obtained using the BBR.

7.6 Recommendation for Future Work

Based on the work described in this report, the following recommendations are made.

It is recommended that AASHTO TP125-16: Standard Method of Test for Determining the Flexural Creep Stiffness of Asphalt Mixtures Using the Bending Beam Rheometer (BBR) be adopted as a method to control the performance of asphalt mixtures at low temperatures.

It is recommended that more research be done regarding methods to reduce the variability of the intermediate-temperature test before it is adopted. This includes determination of alternate, more robust test parameters outside those that are used to calculate the Flexibility Index, evaluation of the effect of sample preparation (compaction and cutting) on the results, alternate geometries such as testing in the indirect tensile mode where no sample cutting is required, and different loading rates.

It is recommended that field performance data be collected on the seven pavement sections evaluated as part of this study. Knowing the actual performance would assist in selecting appropriate loading rates or any other parameter for the test. Knowing the actual field performance will allow development of a threshold that can eventually be used as a specification limit.

REFERENCES

[1] Romero, P., C. H. Ho, and K. VanFrank. Development of Methods to Control Cold Temperature and Fatigue Cracking of Asphalt Mixtures. Report UT-10.08, Utah Department of Transportation, May 2011.

[2] Romero, P., VanFrank, K., Balanced Asphalt Concrete Mix Performance Phase II: Analysis of BBR and SCB Tests, Report UT-17.21. Utah Department of Transportation Research Division, 2017.

[3] Marasteanu, M., Zofka, A., Turos, M., Li, X., Velasquez, R., Li, X., and Ojo, J. Investigation of low temperature cracking in asphalt pavements national pooled fund study 776, 2007.

[4] Jones, Z. Romero, P., and VanFrank, K.: "Development of Low-Temperature Performance Specifications for Asphalt Mixtures using the Bending Beam Rheometer." Journal of Road Materials and Pavement Design, Taylor Francis Volume 15, Issue 3, Pp 574-587

[5] Ouchterlony, F. Fracture toughness testing of rock with core based specimens. Engineering Fracture Mechanics, 35(1),1990, p 351-366.

[6] Chong, K., and Kuruppu, M. D. New specimen for fracture toughness determination for rock and other materials. International Journal of Fracture, 26(2), 1984, R59-R62.

[7] Krans, R. L., Tolman, F., and Van de Ven, M. F. C. Semi-circular bending test: a practical crack growth test using asphalt concrete cores. In RILEM proceedings, Chapman and Hall. 1996, p 123-132.

[8] Hofman, R., Oosterbaan, B., Erkens, S. M. J. G., and van der Kooij, J. Semi-circular bending test to assess the resistance against crack growth. In 6th RILEM Symposium on performance testing and evaluation of bituminous materials, 2003, p 257-263.

[9] Molenaar, A. A. A., Scarpas, A., Liu, X., and Erkens, S. M. J. G. Semi-circular bending test; simple but useful? Journal of the Association of Asphalt Paving Technologists, 71, 2002.

[10] Jahanbakhsh, H., Karimi, M. M., and Tabatabaee, N. Experimental and numerical investigation of low-temperature performance of modified asphalt binders and mixtures. Road Materials and Pavement Design, 2016, 1-22.

[11] Wang, H., Zhang, C., Li, L., You, Z., and Diab, A. Characterization of Low Temperature Crack Resistance of Crumb Rubber Modified Asphalt Mixtures Using Semi-Circular Bending Tests. Journal of Testing and Evaluation, 44(2), 2016, p 847-855.

[12] EN 12697-44:2010: Bituminous Mixtures—Test Methods for Hot Mix Asphalt Part 44: Crack Propagation by Semi-Circular Bending Test, European Committee for Standardization, Brussels, Belgium, 2010.

[13] AASHTO TP 105-13. Standard Method of Test for Determining the Fracture Energy of Asphalt Mixtures Using the Semicircular Bend Geometry (SCB), American Association of State and Highway Transportation Officials, 2013.

[14] Arabani, M., and Ferdowsi, B. Evaluating the semi-circular bending test for HMA mixtures. International Journal of Engineering A: Basics, 22(1), 2009, 47-58.

[15] Huang, B., Shu, X., and Zuo, G. Using notched semi circular bending fatigue test to characterize fracture resistance of asphalt mixtures. Engineering Fracture Mechanics, 109, 2013, p 78-88.

[16] Mull, M. A., Stuart, K., and Yehia, A. Fracture resistance characterization of chemically modified crumb rubber asphalt pavement. Journal of Materials Science, 37(3), 2002, 557-566.

[17] Mull-Aglan, M. A., and Othman, A. Mohammad, L. N., Fatigue crack growth analysis of hot-mix asphalt employing semicircular notched bend specimen. Transportation research board 85th annual meeting (No. 06-1665), 2006.

[18] Mohammad, L. N., Wu, Z., and Aglan, M. A. Characterization of fracture and fatigue resistance on recycled polymer-modified asphalt pavements. In Proceedings of the RILEM: 5th International Conference on Cracking in Pavements Mitigation, Risk Assessment and Prevention, France, 2004, p 375-382.

[19] Mohammad, L. N., Kim, M., and Elseifi, M. Characterization of asphalt mixture's fracture resistance using the semi-circular bending (SCB) test. 7th RILEM international conference on cracking in pavements, Springer Netherlands. 2012, p 1-10.

[20] Al-Qadi, I. L., Ozer, H., Lambros, J., El Khatib, A., Singhvi, P., Khan, T., ... and Doll, B. Testing protocols to ensure performance of high asphalt binder replacement mixes using RAP and RAS. Illinois Center for Transportation/Illinois Department of Transportation, 2015.

[21] Nsengiyumva, G., Kim, Y. R., and You Ph D, T. Development of a Semicircular Bend (SCB) Test Method for Performance Testing of Nebraska Asphalt Mixtures, 2015.

[22] ASTM D8044-16, Standard Test Method for Evaluation of Asphalt Mixture Cracking Resistance using the Semi-Circular Bend Test (SCB) at Intermediate Temperatures, ASTM International, West Conshohocken, PA, 2016.

[23] AASHTO TP 124-16 Provisional Standard Method of Test for Determining the Fracture Potential of Asphalt Mixtures Using Semicircular Bend Geometry (SCB) at Intermediate Temperature, American Association of State and Highway Transportation Officials, 2016.

[24] Li, Y., Romero, P., Sudbury, D., and Allen, C.: "Repeatability and Reproducibility of Low-Temperature Testing of Asphalt Mixtures Using the Bending Beam Rheometer." Proceedings ASCE 16th International Conference on Cold Regions Engineering. Salt Lake City, UT. ISBN (PDF): 978-0-7844-7931-5, Salt Lake City, UT. W.S. Guthrie, ed. Pp 217-228 (July, 2015)

[25] Clendennen, C., and Romero, P.: "Evaluating the Representative Volume Element of Asphalt Concrete Mixture Beams for Testing in the Bending Beam Rheometer." Multi-Scale Modeling and Characterization of Infrastructure Materials, N. Kringos, B. Birgisson, D. Frost, and L. Wang, Eds. Stockholm, Sweden ISBN 978-94-007-6877-2. Pp. 13-30 (10-12 June, 2013)

[26] Salkind, N.J. Statistics: for people who think they hate statistics. 2nd Edition, SAGE Publications Inc. UK, 2007.

[27] Lehman, E. Theory of Point Estimation. Wadsworth and Brooks Cole, CA, 1991.

[28] Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. Educational and Psychological Measurement, 50(4), 717-729.

[29] Seo, S., A review and comparison of methods for detecting outliers in univariate data sets, Doctoral dissertation, University of Pittsburgh, 2006.

[30] ACI (American Concrete Institute). (2011). Guide to evaluation of strength test results of concrete.

[31] Ozer, H., Al-Qadi, I. L., Barber, E., Okte, E., Zhu, Z., & Wu, S., Evaluation of I-FIT Results and Machine Variability using MnRoad Test Track Mixtures. Illinois Center for Transportation/Illinois Department of Transportation, 2017.

APPENDIX A: DATA

All of the data from testing was collected using electronic data acquisition of force, displacement, and temperature sensors. The data was collected in non-proprietary CSV format as generated by the data acquisition system. Spreadsheets were used to summarize and analyze the data. The raw data, called primary data, has been preserved and archived at Zenodo (https://zenodo.org/), an international repository/archive of research outputs from across all fields of research. Zenodo is listed as conforming to the USDOT Public Access Plan (https://ntl.bts.gov/publicaccess/repositories.html). According to Zenodo's policy, data entries remain accessible forever.

The data for the BBR field study described in Section 4 is accessible at the following link: https://doi.org/10.5281/zenodo.2827033

Romero. (2019). Evaluation of Materials for Low Temperature Asphalt Pavement Performance, BBR Field study [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2827033

The data for the SCB laboratory study described in Section 5 is accessible at the following link: http://doi.org/10.5281/zenodo.2565717

Pedro Romero. (2019). Evaluation of Materials for Asphalt Mixture Performance, Semi-Circular Bend Laboratory Tests (Version 1) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2565717

The data for the SCB field study described in Section 6 is accessible at the following link: http://doi.org/10.5281/zenodo.2574116

Pedro Romero. (2019). Evaluation of Materials for Asphalt Mixture Performance, Semi-Circular Bend Field Material (Version 1) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2574116

A README file, including the metadata/information required to repeat the research, is included along with the data in the archive. Zenodo will provide proper citation for users to incorporate the data into their publications and will have a memorandum of understanding (MOU) stating that users may not re-release the data to a third party, but direct them back to the repository.